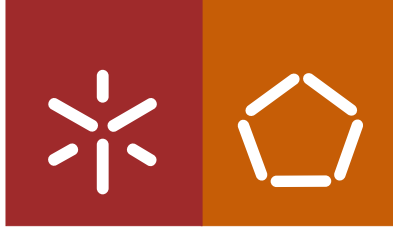


Universidade do Minho
Escola de Engenharia

Sérgio Rafael Mano Pereira

**Deteção automática de Espaços de
Virchow-Robin em imagens de ressonância
magnética**



Universidade do Minho

Escola de Engenharia

Sérgio Rafael Mano Pereira

**Deteção automática de Espaços de
Virchow-Robin em imagens de ressonância
magnética**

Dissertação de Mestrado
Mestrado Integrado em Engenharia Biomédica
Ramo Eletrónica Médica

Trabalho efetuado sob orientação do
Professor Doutor Carlos Alberto Batista da Silva
e do
Professor Doutor Nuno Jorge Carvalho Sousa

outubro de 2013

DECLARAÇÃO

Nome: Sérgio Rafael Mano Pereira

Endereço eletrónico: a55586@alunos.uminho.pt

Número do cartão de cidadão: 13727860

Título da Dissertação: Detecção automática de Espaços de Virchow-Robin em imagens de ressonância magnética

Orientadores: Professor Doutor Carlos Alberto Batista da Silva e Professor Doutor Nuno Jorge Carvalho Sousa

Ano de conclusão: 2013

Designação do Mestrado: Mestrado Integrado em Engenharia Biomédica

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, ___ / ___ / _____

Assinatura: _____

Agradecimentos

Gostaria de começar por agradecer aos meus orientadores, os Professores Doutores Carlos A. Silva e Nuno Sousa. Ao Professor Doutor Carlos A. Silva agradeço as suas orientações, conselhos e disponibilização de recursos que foram fundamentais para esta dissertação, e possibilidade de participação no concurso do *MICCAI Grand Challenge on MR Brain Image Segmentation 2013*. Gostaria de agradecer ao Professor Doutor Nuno Sousa pela possibilidade de fazer dissertação neste interessante tema, e pelo acesso à Escola de Ciências da Saúde e ao Centro Clínico Académico do Hospital de Braga, possibilitando-me o acesso à base de dados de imagens de ressonância magnética.

Não posso passar sem agradecer ao Dr. José António Mariz e ao Eng. Paulo Marques, da Escola de Ciências da Saúde, que sempre se mostraram disponíveis para discutir este trabalho e a base de dados utilizada para os Espaços de Virchow-Robin.

Também gostaria de agradecer ao Adriano e ao Hugo pela sua contribuição e ajuda.

Pelo excelente ambiente de trabalho e boa disposição, aqui deixo o meu agradecimento às pessoas do laboratório.

Agradeço muito aos meus amigos, por todo o apoio que deles recebi. Aqui destaco os meus amigos do curso, com os quais convivi durante esta etapa que durou cinco anos. De facto, é um privilégio tê-los encontrado. Gostaria de realçar, sem detrimento para com os restantes, a Joana, o José e o Pedro, pela convivência quase diária durante este ano e o anterior, companheirismo e amizade.

Por fim, porém não com reconhecimento menor, agradeço aos meus pais e irmão todo o suporte, apoio, motivação e carinho, não só durante este ano de dissertação, mas sempre.

A todas estas pessoas,

Muito Obrigado.

Resumo

Os Espaços de Virchow-Robin estão compreendidos entre as paredes das artérias perforantes do encéfalo e meninges, estando preenchidos por líquido. Por vezes ficam dilatados, tornando-se visíveis em imagens de ressonância magnética, apesar das suas dimensões próximas da resolução dos *scanners* atuais. Ao longo dos anos tem sido colocada a hipótese de o número dessas estruturas dilatadas se correlacionarem com algumas doenças, por exemplo acidentes vasculares cerebrais ou demência. No entanto, estes estudos baseiam-se em contagens semi-quantitativas em algumas regiões de interesse, por visualização das imagens.

Assim, pretendeu-se implementar um sistema automático que, utilizando imagens de ressonância magnética, fosse capaz de extrair regiões de interesse, nomeadamente os gânglios da base e a substância branca, e fazer a deteção e contagem de Espaços de Virchow-Robin.

Nesse sentido construiu-se uma *pipeline* para o pré-processamento, englobando a correção da falta de homogeneidade, remoção do crânio, do ruído e normalização das intensidades. No pré-processamento e na extração dos gânglios da base utilizou-se *software* existente, mas para a segmentação da substância branca desenvolveu-se um algoritmo que utiliza *Random Decision Forests*. Já para a deteção dos Espaços de Virchow-Robin implementou-se um algoritmo que modela as suas propriedades com *Marked Point Process*, e procura a configuração que melhor se adequa através da otimização com *Reversible Jump Markov Chain Monte Carlo* e *simulated annealing*.

O algoritmo de extração de substância branca também demonstrou resultados positivos na segmentação de substância cinzenta, porém o método de remoção do crânio não excluiu alguns tecidos, fazendo com que os resultados da segmentação do líquido cefalorraquidiano fossem piores. Foi validado na base de dados MRBrainS, demonstrando robustez relativamente à presença de lesões da substância branca, desde que existisse a sequência FLAIR.

Por seu lado, o algoritmo de deteção de Espaços de Virchow-Robin foi aplicado em imagens MPRAGE com resolução isotrópica de 1 mm. Apesar de não ter sido possível validar, observou-se que a sua performance foi superior na substância branca do que nos gânglios da base, devendo, no futuro, desenvolver-se filtros mais adequados para a segunda região de interesse. Também se desenvolveu uma aplicação para visualização das estruturas detetadas, e da sua distribuição espacial.

Abstract

Virchow-Robin Spaces surround the walls of the perforating arteries of the brain, being bounded by meninges and filled with cerebrospinal fluid. Sometimes they get dilated, becoming visible in magnetic resonance images, although their dimensions are near the current scanners' resolution. Over the years, the hypothesis that the number of these dilated structures may be correlated with some diseases has been studied, for example with strokes or dementia. However, these studies are based on semi-quantitative counts in some regions of interest, by visualization of the images.

Therefore, it was intended to implement an automatic system that, using magnetic resonance images, was able to extract regions of interest, namely the basal ganglia and the white matter, and detect and count dilated Virchow-Robin Spaces.

In order to do so, it was built a pipeline for the pre-processing of the images, which included the inhomogeneity correction, skull stripping, denoising and intensities normalization. For the pre-processing procedures and extraction of the basal ganglia it was used already existent software, but for the segmentation of the white matter it was developed an algorithm that employs Random Decision Forests. For the detection of the Virchow-Robin Spaces it was implemented an algorithm that models their properties with a Marked Point Process, and searches for the best configuration of these structures among the candidates by optimization of the model with Reversible Jump Markov Chain Monte Carlo and simulated annealing.

The white matter segmentation algorithm also demonstrated positive results in the segmentation of grey matter, but the skull stripping method wasn't able to exclude some tissues, resulting in worse performance for the cerebrospinal fluid. It was validated in the MRBrainS database, demonstrating that it is robust in the presence of white matter lesions, if the FLAIR sequence is available.

On the other hand, the algorithm for detecting dilated Virchow-Robin Spaces was applied in MPRAGE sequences acquired with isotropic resolution of 1 mm. Although it was not possible to validate, it was observed that the algorithm's performance was superior in the white matter than in the basal ganglia, so, in the future, better filters for the second region of interest should be developed. It was, also, built an application to visualize the detected structures and their spatial distribution.

Índice

Agradecimentos	iii
Resumo	v
Abstract.....	vii
Índice	ix
Índice de Figuras.....	xiii
Índice de Tabelas	xvii
Lista de Acrónimos	xix
Capítulo 1 Introdução.....	1
1.1. Motivação	1
1.2. Objetivos.....	2
1.3. Contribuições	3
1.4. Organização da dissertação.....	4
Capítulo 2 Contexto clínico	5
2.1. Imagiologia por ressonância magnética	5
2.2. O encéfalo	8
2.2.1. Proteção do encéfalo.....	9
2.2.1.1. Meninges	10
2.2.1.2. Líquido cefalorraquidiano.....	10
2.2.1.3. Barreira hematoencefálica	10
2.2.2. Telencéfalo.....	11
2.2.2.1. Córtex cerebral	11
2.2.2.2. Substância branca cerebral	12
2.2.2.3. Gânglios da base	12
2.2.3. Importância da segmentação de tecidos do encéfalo	12
2.2.4. Identificação dos tecidos do encéfalo em imagens de ressonância magnética	13
2.3. Espaços de Virchow-Robin	14
2.3.1. Dilatação dos Espaços de Virchow-Robin.....	15

2.3.2. Identificação de Espaços de Virchow-Robin em imagens de ressonância magnética e localizações mais comuns	16
2.3.3. Espaços de Virchow-Robin como biomarcadores de algumas doenças.....	17
2.3.3.1. Doença dos pequenos vasos cerebrais	18
2.3.3.2. Problemas cognitivos	19
2.3.3.3. Outras relações.....	19
2.3.4. Métodos de contagem dos Espaços de Virchow-Robin dilatados.....	20
2.4. Sumário.....	21
Capítulo 3 Estado da arte.....	23
3.1. Segmentação de tecidos do encéfalo.....	23
3.2. Segmentação e identificação de Espaços de Virchow-Robin dilatados	29
3.3. Sumário.....	30
Capítulo 4 Modelização através de <i>Point Process</i>.....	33
4.1. Point Process.....	33
4.1.1. Poisson Point Process	35
4.1.2. Gibbs Point Process	35
4.2. Marked Point Process.....	37
4.2.1. Gibbs Point Process adaptado para a extração de segmentos de reta	38
4.3. Sumário.....	39
Capítulo 5 Otimização através de métodos de Markov Chain Monte Carlo	41
5.1. Monte Carlo.....	42
5.1.1. Amostragem de rejeição.....	42
5.2. Markov Chain Monte Carlo	43
5.2.1. Cadeias de Markov.....	43
5.2.2. Algoritmo de Metropolis-Hastings	45
5.2.2.1. Distribuições propostas.....	47
5.2.2.2. Demonstração da condição de reversibilidade do algoritmo de Metropolis-Hastings.....	47
5.2.3. <i>Reversible Jump Markov Chain Monte Carlo</i>	49
5.3. <i>Simulated annealing</i> como forma de otimização.....	51
5.4. Sumário.....	54

Capítulo 6 Reconhecimento de padrões	57
6.1. Conceitos globais de reconhecimento de padrões	58
6.1.1. Classificador	58
6.1.2. Características.....	58
6.1.3. Avaliação do modelo de classificação.....	59
6.2. Fundamentos de <i>Random Forests</i> enquanto classificador	60
6.2.1. <i>Decision Trees</i>	61
6.2.1.1. Modelo matemático das <i>Decision Trees</i>	62
6.2.1.2. Treino e teste das <i>Decision Trees</i>	62
6.2.2. <i>Random Forests</i>	65
6.2.2.1. Efeito dos parâmetros das <i>Random Forests</i>	66
6.2.2.2. Propriedades de margem máxima	67
6.3. Sumário.....	69
Capítulo 7 Trabalho experimental	71
7.1. Pré-processamento.....	71
7.1.1. Correção da falta de homogeneidade	73
7.1.1.1. Método N3 e N4	74
7.1.2. Remoção do crânio	77
7.1.3. Normalização das intensidades	78
7.1.3.1. Transformação do histograma	79
7.1.4. Remoção de ruído	81
7.1.4.1. Prefiltered Rotationally Invariant Non-Local Mean Filter.....	81
7.2. Extração da Região de Interesse	83
7.2.1. Segmentação subcortical	83
7.2.2. Segmentação de tecidos do encéfalo e extração da substância branca.....	85
7.2.2.1. Extração de características	87
7.2.2.2. Classificação e pós-processamento.....	89
7.3. Detecção dos Espaços de Virchow-Robin.....	91
7.3.1. Filtragem	91
7.3.2. <i>Marked Point Process</i>	93
7.3.3. Otimização com RJMCMC	97
7.4. Aplicação para visualização de Espaços de Virchow-Robin	100

7.5. Sumário.....	102
Capítulo 8 Resultados e discussão	103
8.1. Base de dados MRBrainS	103
8.2. Base de dados para detecção de Espaços de Virchow-Robin.....	103
8.3. Métricas de avaliação	104
8.4. Avaliação da segmentação de tecidos do encéfalo	104
8.4.1. Avaliação do método <i>Atropos</i>	105
8.4.2. Avaliação individual das características na classificação.....	106
8.4.3. Avaliação conjunta das características na classificação.....	109
8.4.4. Avaliação da remoção de ruído	111
8.4.5. Avaliação do número de pontos e classes de treino	112
8.4.6. Avaliação do critério de divisão dos nós	113
8.4.7. Avaliação do número de <i>Decision Trees</i> e profundidade máxima dos nós.....	114
8.4.8. Avaliação do efeito do pós-processamento	116
8.4.9. Avaliação no conjunto de teste	118
8.4.10. Avaliação da segmentação utilizando apenas a sequência T1	120
8.4.11. Avaliação da segmentação de sequências MPRAGE.....	122
8.5. Detecção automática de Espaços de Virchow-Robin	124
8.5.1. Simulação de Espaços de Virchow-Robin.....	124
8.5.2. Detecção em imagens reais	126
8.5.2.1. Detecção de Espaços de Virchow-Robin nos gânglios da base.....	126
8.5.2.2. Detecção de Espaços de Virchow-Robin na substância branca	129
8.6. Sumário.....	133
Capítulo 9 Conclusão e trabalho futuro	135
9.1. Conclusões.....	135
9.2. Trabalho futuro.....	136
Bibliografia.....	139

Índice de Figuras

Figura 2.1 - Figura ilustrativa da física e fundamentos da ressonância magnética.	7
Figura 2.2 - Encéfalo, e respetivas regiões.....	9
Figura 2.3 - Lobos do córtex cerebral.....	11
Figura 2.4 - Corte axial no encéfalo, onde se pode observar os gânglios da base, o córtex cerebral e a substância branca cerebral.	12
Figura 2.5 - <i>Slice</i> axial de uma imagem T1w (esquerda) e <i>slice</i> coronal de uma imagem T2w. .	14
Figura 2.6 – Espaços de Virchow-Robin.	14
Figura 2.7 - Representação esquemática dos Espaços de Virchow-Robin nos gânglios da base e no córtex.	15
Figura 2.8 - Exemplos de Espaços de Virchow-Robin nos três tipos de locais mais comuns.	17
Figura 4.1 - Exemplo de <i>Point Process</i>	34
Figura 4.2 – <i>Marked Point Process</i> com marcas qualitativas (cor) e marcas quantitativas (raio)..	37
Figura 5.1 - Algoritmo para o método Monte Carlo de amostragem de rejeição.	43
Figura 5.2 - Algoritmo de Metropolis-Hastings.....	46
Figura 5.3 - Demonstração da influência do número de iterações na simulação.....	46
Figura 5.4 - Teste do efeito do desvio padrão da distribuição gaussiana proposta durante a simulação.	48
Figura 5.5 - Exemplificação do princípio do método <i>Reversible Jump Markov Chain Monte Carlo</i>	49
Figura 5.6 - Algoritmo <i>Reversible Jump Markov Chain Monte Carlo</i>	51
Figura 5.7 - Algoritmo de Metropolis-Hastings, alterado para fazer otimização global, através de <i>simulated annealing</i>	53
Figura 5.8 - Efeito da taxa de arrefecimento no método de <i>simulated annealing</i>	54
Figura 6.1 – <i>Decision Tree</i>	61
Figura 6.2 - Demonstração da divisão dos pontos de duas maneiras diferentes num nó de um <i>Decision Tree</i>	64
Figura 6.3 – Efeito de alguns parâmetros das <i>Random Forests</i> na divisão dos pontos de treino.	67
Figura 6.4 – Propriedade de margem máxima das <i>Random Forests</i>	68

Figura 7.1 - Diagrama de blocos representativo da <i>pipeline</i> do método desenvolvido, evidenciando as etapas principais.....	72
Figura 7.2 – Efeito da correção da falta de homogeneidade com N3 e N4.....	76
Figura 7.3 - Correção da falta de homogeneidade com o método N4	77
Figura 7.4 – Remoção do crânio.....	78
Figura 7.5 – Normalização das intensidades	80
Figura 7.6 - Efeito da remoção de ruído, num plano axial	82
Figura 7.7 - Extração dos gânglios da base e região subinsular	85
Figura 7.8 - <i>Pipeline</i> do algoritmo de segmentação de tecidos, e extração da substância branca.	86
Figura 7.9 - Algumas características	90
Figura 7.10 - <i>Pipeline</i> para a deteção automática de Espaços de Virchow-Robin.	91
Figura 7.11 – Filtros de propriedades dos Espaços de Virchow-Robin.	92
Figura 7.12 - Tipos de interação entre Espaços de Virchow-Robin.....	95
Figura 7.13 - Função para a avaliação da intensidade dos <i>voxels</i> e da vizinhança $\mathcal{N}1$ imediatamente a seguir aos extremos dos candidatos a Espaços de Virchow-Robin	96
Figura 7.14 - Movimento <i>split</i>	99
Figura 7.15 - Janela principal da aplicação para visualização dos Espaços de Virchow-Robin na região de interesse.....	101
Figura 7.16 - Painel 3D da aplicação	101
Figura 7.17 - Painel <i>Overlay_3D</i>	102
Figura 8.1 - Avaliação do algoritmo <i>Atropos</i>	105
Figura 8.2 - <i>Slice</i> axial segmentado com o <i>Atropos</i>	106
Figura 8.3 - Avaliação individual das características sobre a segmentação, quando utilizadas em conjunto com as intensidades das sequências.....	107
Figura 8.4 - Segmentação resultante de cada característica, em conjunto com as intensidades	109
Figura 8.5 - Avaliação conjunta das características, por classe, em média	110
Figura 8.6 - Exemplos de segmentação, variando o número de pontos e de classes de treino.	113
Figura 8.7 - Avaliação do número de <i>Decision Trees</i> e profundidade máxima, para o conjunto das três classes.....	115
Figura 8.8 - Efeito do pós-processamento	117

Figura 8.9 - Impacto da remoção do crânio sobre o líquido cefalorraquidiano	119
Figura 8.10 - Segmentação de um sujeito do conjunto de teste.....	119
Figura 8.11 - Avaliação do número de <i>Decision Trees</i> e da profundidade máxima, para o conjunto das três classes, utilizando apenas a sequência T1 como base.	121
Figura 8.12 - Exemplo de segmentação utilizando apenas T1.....	122
Figura 8.13 - Segmentação de tecidos em sequências MPRAGE.	124
Figura 8.14 - Simulação de Espaços de Virchow-Robin.....	125
Figura 8.15 - Energia do modelo <i>Marked Point Process</i> para Espaços de Virchow-Robin nos gânglios da base.....	127
Figura 8.16 - Quantidade de Espaços de Virchow-Robin aceites por comprimento, nos gânglios da base.....	128
Figura 8.17 - Espaços de Virchow-Robin nos gânglios da base	129
Figura 8.18 - Espaços de Virchow-Robin detetados nos gânglios da base	130
Figura 8.19 - Energia do modelo <i>Marked Point Process</i> para Espaços de Virchow-Robin na substância branca cerebral.....	131
Figura 8.20 - Quantidade de Espaços de Virchow-Robin aceites por comprimento, na substância branca cerebral.....	131
Figura 8.21 - Espaços de Virchow-Robin na substância branca.....	132
Figura 8.22 - Espaços de Virchow-Robin detetados na substância branca cerebral.....	133

Índice de Tabelas

Tabela 2.1 - Intensidade apresentada por alguns tecidos em imagens <i>T1-weighted</i> e <i>T2-weighted</i>	8
Tabela 2.2 - Sumário das estruturas pertencentes a cada região do encéfalo, e respectivas funções	9
Tabela 2.3 - Métodos semi-quantitativos de contagem dos Espaços de Virchow-Robin dilatados.	20
Tabela 2.4 - Métodos semi-quantitativos de contagem dos Espaços de Virchow-Robin dilatados (continuação)	21
Tabela 3.1 - Estado da arte de segmentação de tecidos	31
Tabela 3.2 - Estado da arte de algoritmos de detecção de Espaços de Virchow-Robin.....	32
Tabela 7.1 - Comparação entre diferentes implementações de <i>Random Forests</i>	90
Tabela 7.2 - Parâmetros do modelo <i>Marked Point Process</i> para Espaços de Virchow-Robin.....	97
Tabela 8.1 - Avaliação de cada característica individualmente, em conjunto com as intensidades.	107
Tabela 8.2 - Avaliação conjunta das características	110
Tabela 8.3 - Avaliação da remoção do ruído na segmentação.....	111
Tabela 8.4 - Avaliação do número de pontos e classes para treinar.....	112
Tabela 8.5 - Avaliação do critério de divisão nos nós	114
Tabela 8.6 - Avaliação do número de <i>Decision Trees</i> e da profundidade máxima dos nós, para cada classe.....	115
Tabela 8.7 - Avaliação do pós-processamento para a correção de pequenas regiões isoladas, testando-se o número de pontos máximos de cada região isolada.....	117
Tabela 8.8 - Resultados da segmentação no conjunto de teste.	118
Tabela 8.9 - Avaliação do número de <i>Decision Trees</i> e da profundidade máxima dos nós, para cada classe.....	120
Tabela 8.10 - Número de candidatos a EVR e segmentos aceites nos gânglios da base, para cada percentagem de remoção de pontos.....	128
Tabela 8.11 - Número de candidatos a Espaço de Virchow-Robin e segmentos aceites na substância branca cerebral cerebral, para cada percentagem de remoção de pontos.	131

Lista de Acrónimos

3D – Tridimensional

ANTs - *Advanced Normalization Tools*

AVC – Acidente vascular cerebral

BET – *Brain Extranction Tool*

DT – *Decision Tree*

EM – *Expectation-Maximization*

EVR – Espaços de Virchow-Robin

FAST - *FMRIB's Automated Segmentation Tool*

FSL - *FMRIB Software Library*

IBSR – *Internet Brain Segmentation Repository*

IR – *Inversion recovery*

LCR – Líquido cefalorraquidiano

MCMC – *Markov Chain Monte Carlo*

MH – *Metropolis-Hastings*

MPP – *Marked Point Process*

MRF – *Markov Random Field*

N3 - *Nonparametric Nonuniform Intensity Normalization*

N4 – *Nick's Nonparametric Nonuniform Intensity Normalization*

PP – *Point Process*

PRINLM - *Prefiltered Rotationally Invariant Non-Local Mean*

RF – *Random Forest*

RJMCMC – *Reversible Jump Markov Chain Monte Carlo*

RM – Ressonância magnética

SB – Substância branca

SC – Substância cinzenta

T1w – *T1-weighted*

T2w – *T2-weighted*

TE – Tempo de eco

TR – Tempo de repetição

Capítulo 1 Introdução

Neste capítulo irá começar-se por expor a motivação para a deteção automática de Espaços de Vichow-Robin; depois, irá descrever-se os objetivos, as contribuições e a organização da presente dissertação.

1.1. Motivação

O encéfalo faz parte do sistema nervoso central, sendo constituído maioritariamente por substância cinzenta (SC) e substância branca (SB), encontrando-se rodeado, e preenchido nos seus sulcos e ventrículos, por líquido cefalorraquidiano (LCR) [1]. Ao redor das artérias perfurantes que atravessam este órgão existe uma cavidade delimitada por meninges, preenchida por líquido, denominada Espaço de Virchow-Robin (EVR) [2].

Por vezes, essas cavidades ficam dilatadas, tornando-se visíveis em imagens de ressonância magnética (RM). Estas estruturas podem ser observadas em qualquer faixa etária e em sujeitos saudáveis, mas tornam-se mais comuns à medida que a idade avança. Todavia, existem estudos que encontraram correlações entre o número de EVR dilatados e alguns problemas de saúde, nomeadamente com problemas dos pequenos vasos cerebrais, por exemplo acidentes vasculares cerebrais (AVCs) lacunares [3], e com demências [4]. Nestes casos, a correlação existia quando os espaços eram contados em regiões específicas do encéfalo, como os gânglios da base, região subinsular e SB cerebral (ou centro semioval) [3], [4]. Assim, tem sido estudada a possibilidade de os EVR dilatados serem um novo biomarcador desses problemas, que ajudem a identificá-los e quantificá-los [4].

Os AVCs lacunares são um tipo de AVC isquémico que resultam da oclusão de uma única artéria perfurante do encéfalo, tendo como consequências lesões de dimensões reduzidas (< 15 mm), e podem ser sintomáticos ou silenciosos. Estima-se que 25% dos AVCs isquémicos sejam deste tipo, e que o primeiro acontecimento ocorra em cerca de 26,3/100.000 pessoas, anualmente. A taxa de mortalidade é 2,5% nos 30 dias seguintes, sendo inferior à dos AVCs não lacunares (cerca de 26%), porém a partir do terceiro ano esta taxa iguala-se em ambos os tipos,

tal como a taxa de recorrência. A percentagem destes pacientes que vem a sofrer de demência vascular nos dois a três anos seguintes é elevada, situando-se nos 11% [5], [6].

A demência caracteriza-se pela diminuição das capacidades cognitivas das pessoas, ao ponto de impedir que executem tarefas do quotidiano. A demência vascular é provocada por problemas no fornecimento de sangue no encéfalo, sendo a segunda demência mais comum, ficando atrás apenas da Doença de Alzheimer. Na Europa corresponde a cerca de 17,6% de todas as demências, mas em países asiáticos, como o Japão ou China, este número pode ascender aos 50%. O diagnóstico não é fácil devido à falta de consenso clínico, devendo investigar-se métodos ou biomarcadores que sejam úteis quer seja para o diagnóstico da demência, quer seja para inferir qual a contribuição individual das doenças cerebrovasculares e da Doença de Alzheimer na demência de um paciente que tenha ambas [7].

No conhecimento do autor da dissertação, nenhum trabalho que tenha estabelecido relações entre o número de EVR alargados e problemas de saúde utilizou um sistema automático para a sua deteção e contagem. Nesses estudos foi utilizada observação visual e escalas semi-quantitativas de contagem nas regiões pretendidas, como por exemplo em [3], [4]. Assim, em [8] aponta-se que deveria existir um método de avaliação automático e confiável, capaz de ser aplicado apenas nas regiões que se consideram relevantes para os EVR, que permitisse uniformizar os critérios de identificação dessas estruturas, avaliar a sua real correlação com diversas patologias e eliminar a subjetividade do observador. É neste contexto que se insere a dissertação.

1.2. Objetivos

O principal objetivo da presente dissertação foi a implementação de um sistema automático de deteção de EVR em regiões específicas do encéfalo, como os gânglios da base e a SB cerebral, que permitisse eliminar a subjetividade inerente aos observadores humanos. Para tal, era necessário construir uma *pipeline* de pré-processamento das imagens, uma vez que estas não sofreram qualquer tipo de correção após a aquisição, utilizando *software* disponível; conseqüentemente, foi necessário fazer o levantamento e escolha das ferramentas de neuroimagem mais apropriadas. A possibilidade de se fazer a deteção em determinadas regiões específicas era a principal inovação relativamente aos métodos já existentes, e era importante para futuros estudos de correlação entre o número de EVR nessas regiões e determinadas

condições clínicas, pelo que se deveria pesquisar métodos já disponíveis para as extrair, ou propor uma nova abordagem.

1.3. Contribuições

Ao longo desta dissertação foram desenvolvidos alguns métodos que, no conhecimento do autor, podem ser considerados contribuições originais, nomeadamente:

1. na deteção automática de EVR propôs-se uma *pipeline* de processamento totalmente automática e específica. Porém, a maior contribuição foi a possibilidade de extrair duas regiões de interesse nas quais se faz a deteção de EVR, o que confere maior significado às contagens, não existindo nenhum método na literatura que o tenha feito (em [8], um artigo de revisão acerca da avaliação automática de EVR, os autores consideraram que seria relevante a existência de um método que fosse capaz de restringir a região de contagem).
2. O método de segmentação de SC, SB e LCR com *Random Decision Forests*. As principais contribuições foram a inclusão, como características, de probabilidades *a posteriori* de tecidos, resultantes de uma classificação prévia com um método estatístico, que contava com um modelo *Markov Random Fields*, e estatísticas de vizinhança do *voxel*. Não se utilizou, como característica, informação de locais distanciados do *voxel* como em outros trabalhos que utilizaram este tipo de classificador, como em [9] e [10], e que empregam muitas características (2.000 em [10]). Assim, observou-se que com o conjunto de características utilizadas foi possível obter resultados positivos, com menor carga computacional.
 - a. Este método foi utilizado para participar no *MICCAI Grand Challenge on MR Brain Image Segmentation 2013*, classificando-se em 8/11. O *proceeding* submetido pode ser encontrado no separador de resultados do *website* do concurso¹.
 - b. Está, ainda, a ser preparado um artigo em coautoria com os organizadores e restantes participantes do concurso para ser submetido a uma revista de elevado fator de impacto.

¹ <http://mrbrains13.isi.uu.nl/results.php>, consultado em 30/10/2013

1.4. Organização da dissertação

A dissertação foi dividida em nove capítulos, ao longo dos quais se introduz os fundamentos teóricos, o trabalho experimental, resultados e conclusões. Assim sendo, no capítulo 2 aborda-se o contexto clínico, descrevendo-se o encéfalo, alguns dos seus tecidos e distribuição espacial. Em seguida, introduz-se os EVR, e de que forma a sua dilatação pode estar correlacionada com algumas condições de saúde. No capítulo 3 encontra-se o estado da arte de segmentação de tecidos do encéfalo e deteção de EVR, em imagens de RM.

Do capítulo 4 ao 6 abordam-se os conceitos teóricos. Assim, começa-se por no capítulo 4 expor os conceitos teóricos de modelização por *Point Process*, destacando-se os casos em que existe interações entre pontos. O capítulo 5 é dedicado aos algoritmos *Markov Chain Monte Carlo*, culminando com o *Reversible Jump Markov Chain Monte Carlo* e a utilização destes algoritmos para se fazer otimizações quando utilizados em conjunto com *simulated annealing*. No capítulo 6 abordam-se alguns conceitos de reconhecimentos de padrões e de *Random Decision Forests*, enquanto classificador.

No capítulo 7 descreve-se o trabalho experimental realizado, a *pipeline* construída e as principais opções tomadas. No capítulo 8 reportam-se os resultados obtidos, quer para a segmentação de SC, SB e LCR, quer para a deteção de EVR. Por fim, no capítulo 9 encontram-se as principais conclusões e possível trabalho futuro.

Capítulo 2 Contexto clínico

A utilização da RM para efeitos de diagnóstico e avaliação de tecidos tem crescido exponencialmente, devido ao elevado detalhe das suas imagens. Além disso, é capaz de ser utilizada para fazer aquisições tridimensionais (3D), ou seja, consegue-se avaliar um volume [11]. Assim, é fundamental quando se pretende visualizar estruturas de reduzida dimensão, quase ao nível da resolução da imagem, e com formas 3D conhecidas e relevantes para os distinguir de outras estruturas, como é o caso dos EVR [2].

Neste capítulo irá apresentar-se a técnica de RM, e alguns dos seus conceitos básicos. Em seguida avançar-se-á com a descrição do encéfalo, de modo a facilitar a localização e a familiarização com algumas estruturas anatómicas, e as suas funções. Por fim, irá abordar-se os EVR, descrevendo-se a sua estrutura, o caso em que ficam dilatados, a relação destes últimos com algumas doenças, ou condições, e a forma de identificação.

2.1. Imagiologia por ressonância magnética

Os átomos são constituídos por prótons e neutrões, no seu núcleo, e eletrões em torno deste. Cada um desses constituintes roda sobre um eixo, gerando um momento magnético denominado *spin*. O somatório dos *spins* de cada constituinte permite calcular o *spin* resultante do elemento; tendo em conta que se tratam de cargas elétricas em movimento, então existe um campo magnético associado. Os elementos mais importantes para esta técnica de RM são aqueles cujo número atómico é ímpar, uma vez que se o número de prótons e neutrões for o mesmo os *spins* cancelam-se. Destes, o mais utilizado é o hidrogénio, pois está presente na água e gordura, sendo estes componentes muitos abundantes no organismo humano [12].

Quando se aplica um campo externo, B_0 , a existência do *spin* faz com que o elemento precesse em redor do eixo desse campo a uma frequência (em MHz) que é dependente do elemento, podendo ser calculada pela equação de Larmor (equação (2.1)).

$$\omega_0 = B_0 \times \gamma \quad (2.1)$$

onde ω_0 é a frequência de precessão (MHz), B_0 é o campo magnético externo (T) e γ é o rácio giromagnético do elemento (MHz/T) [12].

Assim, B_0 é um campo magnético forte que é aplicado no material que se quer estudar, o que faz com que o eixo de *spin* dos elementos fique alinhado com o vetor de campo magnético de B_0 (Figura 2.1a). Se, em seguida, se aplicar um pulso de radiofrequência de elevada energia e frequência igual, ou próxima, da frequência de Larmor do elemento cria-se um novo campo magnético, B_1 , que provoca o afastamento do vetor resultante da magnetização do elemento em relação ao vetor de B_0 de um determinado ângulo, e faz com que os *spins* fiquem em fase. Neste momento, passa a existir uma componente de magnetização transversal, à custa da diminuição da componente longitudinal (que tem a direção de B_0). Quando o pulso termina, o vetor de magnetização do elemento volta ao estado de repouso, isto é, volta a ficar alinhado com o vetor de B_0 , ao mesmo tempo que os *spins* voltam a ficar desfasados. Esta mudança da magnetização do elemento cria uma corrente na bobina de receção, que está alinhada com B_0 , cuja amplitude vai diminuindo à medida que a componente transversal também diminui, e é esta corrente que gera o sinal de RM, que é estudado em frequência, após o cálculo da sua transformada de Fourier (Figura 2.1b) [11], [12].

O tempo que o vetor de magnetização dos núcleos atômicos demora a regressar à posição de repouso, ou seja, até se alinharem com B_0 chama-se tempo de relaxação, sendo este importante para a geração de imagem. Portanto, ao tempo que os *spins* demoram até atingir 63% da magnetização longitudinal total chama-se tempo de relaxação longitudinal, através do processo de recuperação T1. Enquanto isso acontece, vai havendo perda de fase entre os *spins*, devido à interação entre partículas vizinhas, o que se traduz na perda de magnetização transversal, que quando perde 69% do seu valor original corresponde ao tempo de relaxação transversal, num processo de decaimento T2 [12].

Dois parâmetros importantes para o contraste da imagem são o tempo de repetição (TR) e o tempo de eco (TE). O primeiro corresponde à diferença temporal entre a aplicação do pulso de radiofrequência e o início do pulso seguinte. Já o TE corresponde ao tempo que decorre entre a aplicação do pulso de radiofrequência e o pico do eco detetado.

O TR relaciona-se fortemente com o contraste em T1, uma vez que, se este for muito longo, pode não existir muitas diferenças entre os sinais dos tecidos, por outro lado, TR curtos, para além de melhorarem o contraste, também podem servir para suprimir determinados tecidos, dada a maior diferença entre os sinais (Figura 2.1b). Por exemplo, é possível suprimir a gordura, pois a sua velocidade de relaxação é maior [11], [12].

Por outro lado, o TE está mais relacionado com T2, afetando o contraste entre tecidos nessa medição. Assim sendo, como se está a medir o decaimento da magnetização dos tecidos, TE maiores favorecem o contraste entre os mesmos, pois a magnetização é sensivelmente igual no início e só se verificam diferenças à medida que o tempo passa (Figura 2.1b). Neste caso, em sólidos o T2 é muito curto, enquanto nos líquidos é longo, o que faz com que tenham intensidade elevada nas imagens [11], [12].

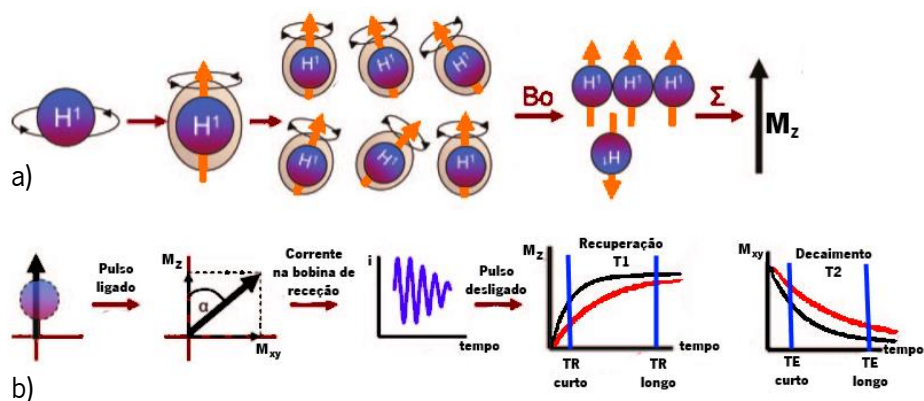


Figura 2.1 - Figura ilustrativa da física e fundamentos da ressonância magnética. Em a) observa-se o efeito da aplicação do campo magnético externo (B_0) sobre os átomos de hidrogênio. Assim, cada átomo tem o seu próprio campo magnético (seta laranja) cuja orientação é aleatória, mas quando se aplica B_0 , verifica-se que os campos magnéticos ficam com a mesma direção do campo externo, e o somatório do campo magnético de todos os átomos resulta num campo magnético resultante não nulo, e alinhado com B_0 . Em b) observa-se o efeito da aplicação do pulso de radiofrequência. Portanto, quando se liga o pulso, o campo magnético resultante dos átomos passa a ter componente transversal (M_{xy}), que cria corrente na bobina de recepção, resultando no sinal de RM. Quando se desliga o pulso pode-se medir o tempo de recuperação T1 e decaimento T2, durante a relaxação. Adaptado de [11].

Existem sequências de pulso que permitem suprimir o sinal de determinados tecidos, chamadas sequências IR (*inversion recovery*). Neste caso, um pulso inicial inverte o *spin* em 180° , seguindo-se a relaxação, porém como os *spins* não estão em fase, não se obtém sinal. Quando o vetor de magnetização do tecido que se pretende anular atinge o plano transversal a sua magnetização transversal resultante é muito baixa, então aplica-se um novo pulso de 90° , que faz com que o vetor de magnetização de todos os tecidos, com a intensidade atual, seja colocado no plano transversal, e como agora os *spins* estão em fase já existe um componente longitudinal da magnetização que se pode medir. As sequências IR mais utilizadas são a FLAIR, que anula fluidos, e a STIR, que anula o sinal da gordura [11], [12].

Apesar de comumente se chamar T1 e T2 às imagens, o nome mais correto é T1-weighted, ou T1w, e T2-weighted, ou T2w. Isto porque, apesar de T1w e T2w demonstrarem as diferenças nos tempos de relaxação T1 e T2, respetivamente, também são dependentes de outros parâmetros [12].

As imagens T1w são as que melhor representam as estruturas anatómicas, podendo também evidenciar patologias ao utilizar-se contraste. Por seu lado, T2w é mais útil na identificação de patologias, já que os tecidos afetados tendem a conter maior concentração de água, surgindo zonas hiperintensas na imagem [11]. Apesar de T1w e T2w serem as sequências básicas de RM, existem outras, que são escolhidas conforme aquilo que se pretende visualizar. Na Tabela 2.1 pode-se consultar a intensidade apresentada pelos tecidos e ar nas imagens de sequências T1w e T2w.

Tabela 2.1 - Intensidade apresentada por alguns tecidos em imagens T1-weighted e T2-weighted. Adaptado de [11].

Intensidade	T1-weighted	T2-weighted
Escura	Ar, tecidos ricos em minerais (osso cortical), fluxo sanguíneo rápido	Ar, tecidos ricos em minerais (osso cortical), fluxo sanguíneo rápido
Baixa	Tecidos de colagénio, líquidos e tecidos ricos em água (fígado, pâncreas, músculo)	Tecidos de colagénio, tecidos ricos em água (fígado, pâncreas, músculo)
Intermédio	Tecidos proteicos (cistos, tumores)	Gordura, medula óssea gordurosa, tecidos ricos em água (fígado, pâncreas, músculo)
Brilhante	Gordura, medula óssea, agentes de contraste, fluxo sanguíneo lento	Líquidos livres, tecidos proteicos, gordura, medula óssea gordurosa

A RM apresenta diversas vantagens, como o facto de ser tridimensional (3D), não utilizar radiação ionizante, não ter efeitos biológicos conhecidos e elevado contraste entre os diferentes tecidos. Porém, o tempo de aquisição, o custo e a interação com materiais ferromagnéticos implantados nos pacientes constituem as principais desvantagens [12].

2.2. O encéfalo

O sistema nervoso encontra-se dividido em duas partes: o sistema nervoso central e o sistema nervoso periférico. A primeira parte é constituída pelo encéfalo e pela espinal medula, e é responsável por interpretar as sensações provenientes dos sentidos, e tomar decisões (racionalis ou reflexas) que são transmitidas aos seus eferentes, como músculos e glândulas. Por seu lado, o sistema nervoso periférico é constituído por nervos que se estendem a partir do sistema nervoso central, ramificando-se na periferia.

O encéfalo é dividido em várias regiões, cada uma com funções específicas: o telencéfalo, o diencefalo, o tronco encefálico e o cerebelo (Figura 2.2). O conjunto do telencéfalo

e o diencefalo é designado como cérebro. Algumas regiões são constituídas por várias estruturas, que se encontram descritas, sumariamente, na Tabela 2.2 [1].

No encéfalo pode-se identificar zonas com coloração acinzentada, tratando-se da SC, que é composta pelas dendrites e corpos celulares dos neurónios. Os axónios com bainha de mielina apresentam uma coloração mais clara, denominando-se essas zonas por SB [13].

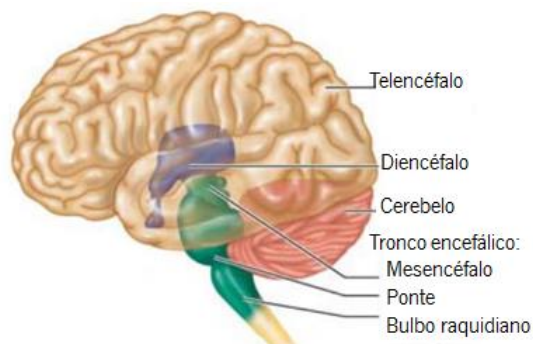


Figura 2.2 - Encéfalo, e respetivas regiões. Adaptado de [1].

Tabela 2.2 - Sumário das estruturas pertencentes a cada região do encéfalo, e respetivas funções. Adaptado de [1].

Região	Estrutura	Função
Telencéfalo	Córtex	Controlo voluntário, interpretação das sensações, atividade muscular, função intelectual e emocional.
	Gânglios da base	Centros motores
Diencefalo	Tálamo	Condução dos impulsos sensoriais e dos centros motores.
	Hipotálamo	Integração do sistema nervoso involuntário, regulação hormonal.
Tronco encefálico	Mesencéfalo	Relação com a visão e audição reflexa. Contém centros motores.
	Ponte	Condução entre os centros encefálicos superiores e inferiores. Condução entre o cérebro e o cerebelo. Controlo da respiração.
	Bulbo raquidiano	Condução entre os centros encefálicos superiores e inferiores. Controlo do ritmo cardíaco, diâmetro dos vasos sanguíneos, ritmo de respiração, tosse, vômitos, entre outros.
Cerebelo		Processa a informação do córtex cerebral motor, equilíbrio e coordenação motora.

2.2.1. Proteção do encéfalo

O encéfalo, tratando-se de um órgão extremamente sensível, encontra-se protegido não só pelo crânio, mas também pelo líquido cefalorraquidiano, pelas meninges e pela barreira hematoencefálica.

2.2.1.1. Meninges

Existem três meninges a envolver o sistema nervoso central, estas conferem proteção quer ao sistema, quer aos vasos sanguíneos, e formam compartimentos que contêm LCR. A meninge mais externa é a duramáter, sendo a mais resistente das membranas. Esta é constituída por duas camadas, a lâmina periosteia e a meníngea, que se encontram juntas na maioria do espaço, e nos locais onde se separam são responsáveis por reunir e direcionar o sangue venoso para as veias jugulares. A segunda meninge é a aracnoide, que é muito fina e elástica, mas não penetra nos sulcos. Entre esta e a duraméter existe o espaço subdural. A meninge mais interna, e mais sensível, é a piamáter. Esta acompanha todos os giros e sulcos do encéfalo e é percorrida por uma rede de pequenos vasos sanguíneos. A separar a piamáter e a aracnoide está o espaço subaracnóideo, preenchido por LCR, sendo também o local por onde passam os maiores vasos sanguíneos que irrigam o encéfalo [1].

2.2.1.2. Líquido cefalorraquidiano

O LCR envolve o encéfalo e a espinal medula, fazendo com que estes órgãos se encontrem a “flutuar”, impedindo que colapsem sobre si mesmos, ao mesmo tempo que os protege contra possíveis choques. Para além da função de suporte e proteção, também contribui para a circulação de alguns compostos, como hormonas, e para a remoção de outros que não são necessários [1].

O líquido é produzido nos plexos coroidais, que se localizam nos ventrículos, os quais estão preenchidos por este líquido, movendo-se destes locais para o espaço subaracnóideo. A sua constituição é semelhante à do plasma sanguíneo, variando um pouco na concentração de iões, contendo também uma menor quantidade de proteínas [1].

2.2.1.3. Barreira hematoencefálica

As células endoteliais dos capilares sanguíneos que irrigam o encéfalo não possuem poros, ao contrário dos outros órgãos, por isso os compostos químicos necessitam de atravessar as células endoteliais, o que faz com que estas funcionem como uma barreira que mantém o ambiente do encéfalo quimicamente estável. É uma barreira seletiva, na medida em que é permeável às substâncias essenciais para o encéfalo, e impermeável às prejudiciais ou não essenciais, sendo capaz de bombear algumas delas para fora do compartimento onde se encontra o referido órgão [14].

2.2.2. Telencéfalo

O telencéfalo é a maior porção do encéfalo, e é o responsável pelas funções mentais superiores, como o pensamento, emoções, e outras tarefas racionais.

Encontra-se dividido em dois hemisférios, direito e esquerdo, unidos pelo corpo caloso, que é o maior trato de axónios. Na sua superfície, o tecido organiza-se na forma de pregas, cujas zonas mais elevadas são denominadas por giros, que se encontram separados entre si por sulcos. Se os sulcos forem profundos, então chamam-se fissuras. De facto, os dois hemisférios são separados pela fissura longitudinal [1], [14].

Cada hemisfério tem três principais regiões: o córtex, a SB interna e os gânglios da base.

2.2.2.1. Córtex cerebral

O córtex é a região mais externa do telencéfalo, constituída por SC. É aqui que ocorrem os pensamentos, emoções e o estado de consciência em geral. Também é responsável pela interpretação das sensações e da comunicação, assim como de qualquer ato voluntário (Figura 2.4) [1], [14].

Para além da grande fissura longitudinal que divide os hemisférios, outras fissuras de grande dimensão dividem o córtex em cinco lobos, cada um com funções específicas. Assim, o lobo frontal é responsável pelo movimento voluntário, personalidade, pensamento e comunicação verbal; o lobo parietal interpreta sensações táteis e a fala, assim como elabora as palavras para expressar pensamentos; no lobo temporal dá-se a interpretação de sensações auditivas e armazenam-se memórias visuais e auditivas; no lobo occipital são interpretadas as sensações visuais e é o local de controlo dos olhos; por fim, a ínsula está envolvida no processo de memória e na resposta autónoma das vísceras relativamente a sensações como olfato, paladar ou dor (Figura 2.3) [14].

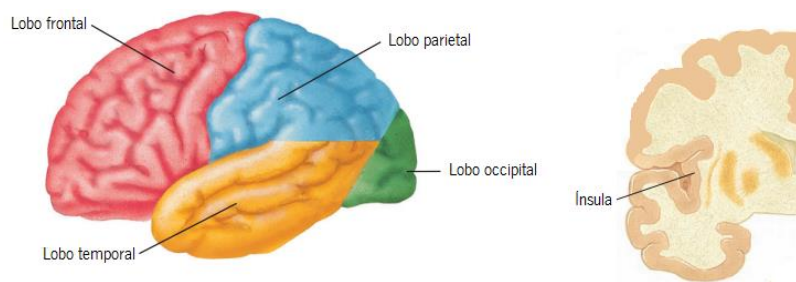


Figura 2.3 - Lobos do córtex cerebral. Adaptado de [14], [15].

2.2.2.2. Substância branca cerebral

É a zona de SB subcortical, que, como já foi referido, é constituída pelos axónios, envolvidos pela bainha de mielina, dos neurónios. Os axónios organizam-se em feixes nervosos que estabelecem a comunicação entre as diversas regiões do cérebro, mas também entre o córtex e as restantes regiões e estruturas do sistema nervoso central (Figura 2.4) [1].

2.2.2.3. Gânglios da base

Os gânglios da base, ou núcleos da base (Figura 2.4), são aglomerados de SC localizados no interior da SB cerebral. São importantes no controlo motor, mais especificamente no início, paragem e monitorização da intensidade dos movimentos comandados pelo córtex, particularmente aqueles que são muito lentos ou estereotipados, como o balanço dos braços durante a marcha [1], [14].

Os núcleos que formam os gânglios da base são o putamen, o núcleo caudado e o globo pálido, um por cada hemisfério. O conjunto formado pelo putamen e globo pálido é denominado por núcleo lentiforme; este e o núcleo caudado formam o corpo estriado. A substância negra e o núcleo subtalâmico estão funcionalmente associados com os referidos núcleos, por isso alguns autores também consideram que pertencem aos gânglios da base [1], [14].

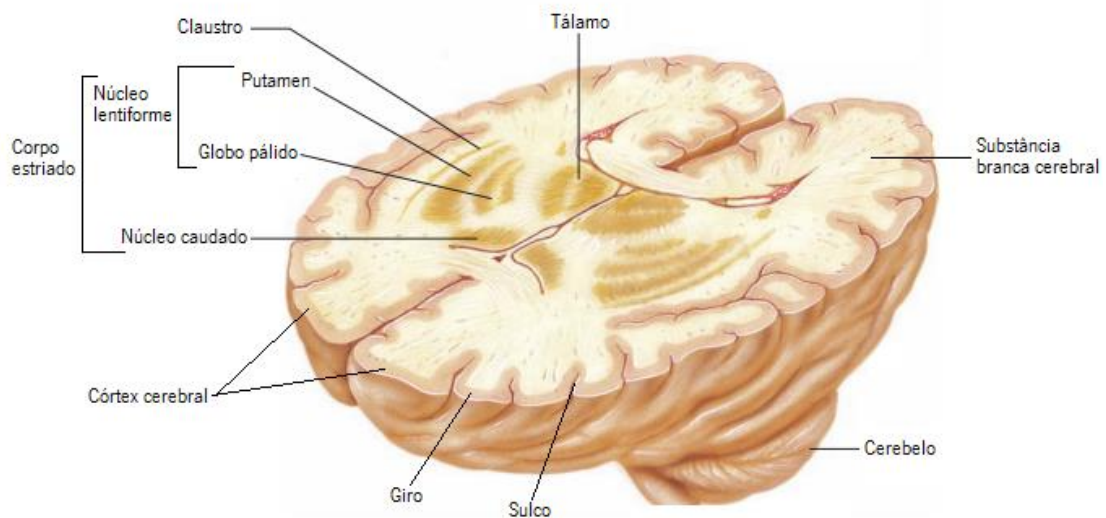


Figura 2.4 - Corte axial no encéfalo, onde se pode observar os gânglios da base, o córtex cerebral e a substância branca cerebral. Adaptado de [14].

2.2.3. Importância da segmentação de tecidos do encéfalo

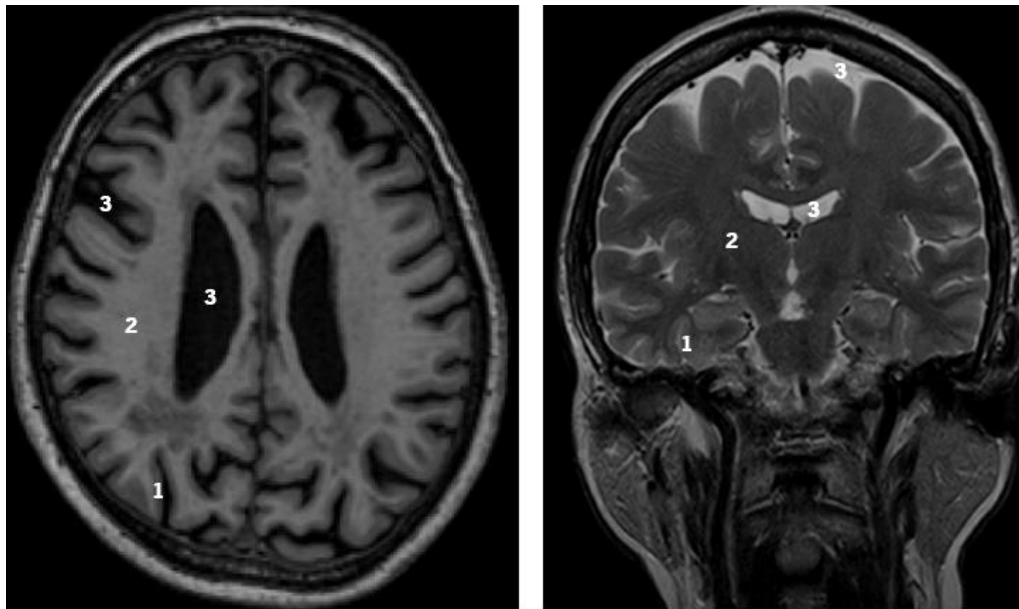
Atualmente, muitos estudos morfométricos utilizam imagens de RM do encéfalo para fazer medições de regiões *in vivo*, quer da deformação, através de mapas de deformação, quer

através da comparação direta dos *voxels*. Esta última é denominada morfometria baseada em *voxels*, e é mais utilizada para avaliar diferenças da SC e SB. Assim, a sua aplicação exige a correção da falta de homogeneidade das imagens, registo para uma imagem de referência e segmentação dos tecidos em SC, SB e LCR, antes de se fazer o estudo estatístico. Entre as vantagens de se utilizarem métodos automáticos apontam-se a reprodutibilidade e a eliminação da subjetividade, inerente ao sujeito que o faria manualmente [16].

A morfometria baseada em *voxels* já foi utilizada em estudos de envelhecimento, concluindo-se que com a idade existe atrofia de SC e SB normal [17]. Já a Doença de Alzheimer está correlacionada com a perda de SC do córtex, particularmente da superfície medial do lobo temporal, que é responsável por algumas memórias [18]. Por seu lado, na Doença de Huntington em estado pré-clínico, ou seja, antes de se observarem sintomas, verificou-se que existe perda no volume de SC, particularmente no corpo estriado e lobo frontal, assim como deterioração da integridade da SB [19]. Também já foram feitos estudos do mesmo género em doentes esquizofrénicos, nos quais se observou défices de volume em várias regiões do cérebro, particularmente em giros específicos [20].

2.2.4. Identificação dos tecidos do encéfalo em imagens de ressonância magnética

A intensidade e o contraste dos tecidos do encéfalo nas imagens de RM dependem de vários parâmetros, como o TR e o TE. No entanto, o tempo de relaxação T1 e T2 são muito elevados para LCR, depois, em níveis mais baixos, segue-se a SC e a SB, por ordem decrescente, mas próximos um do outro [12]. Assim, tendo em conta os conceitos de RM, pode-se esperar que nas imagens T1w o LCR surja muito escuro, enquanto a SC e SB apresentam intensidades intermédias, com a SC mais escura do que a SB. Por outro lado, as imagens T2w são mais dependentes de TE, então o LCR surge muito intenso e a SC e SB apresentam-se com intensidades intermédias, mas com a SC mais clara do que a SB, ao contrário do que acontece em T1w. Na Figura 2.5 pode-se observar um exemplo de uma imagem T1w e T2w do encéfalo.



1 - Substância cinzenta; 2 - Substância branca; 3 - Líquido cefalorraquidiano

Figura 2.5 - *Slice* axial de uma imagem T1w (esquerda) e *slice* coronal de uma imagem T2w. Identificou-se o LCR, SC e SB.

2.3. Espaços de Virchow-Robin

O parênquima cerebral encontra-se percorrido por vasos sanguíneos, responsáveis pela sua irrigação. Ao redor das artérias perfurantes do cérebro estão os EVR, assim denominados em honra dos investigadores que, individualmente, os descreveram [2]. Na Figura 2.6 pode-se observar uma microfotografia de uma artéria envolvida por um EVR, e uma fotografia *post mortem* do encéfalo com EVR.

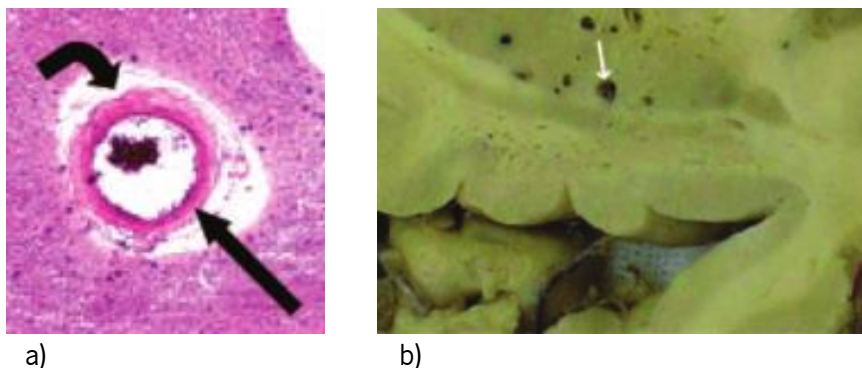


Figura 2.6 – Espaços de Virchow-Robin. Em a) observa-se uma fotografia microscópica de uma artéria perfurante (seta reta) rodeada por um EVR (seta curva). Em b) encontra-se uma secção *post-mortem* do cérebro com a indicação de um EVR (seta branca). Imagens adaptadas de [2], [21], respetivamente.

A piamáter, estando próxima da superfície cerebral estende-se ao longo dos vasos sanguíneos, separando-os quer do espaço subaracnóideo, quer do espaço subpial, delimitando o EVR. Assim sendo, os EVR são a continuação dos espaços perivascularares que rodeiam as arteríolas do espaço subaracnóideo, delimitados por uma leptomeninge de células da piamáter [22]. No entanto, a circunscrição dos EVR junto às artérias perforantes é diferente no córtex e nos gânglios da base. No primeiro caso, eles estão delimitados diretamente pelas artérias perforantes, mas no caso dos gânglios da base existe uma outra leptomeninge mais interna, mais próxima das artérias, tal como se pode observar na Figura 2.7 [2], [22].

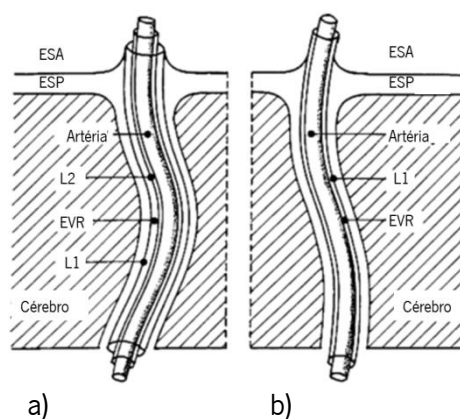


Figura 2.7 - Representação esquemática dos Espaços de Virchow-Robin nos gânglios da base e no córtex. Pode-se observar como a leptomeninge L1 separa o EVR do espaço subpial (ESP) e do espaço subaracnóideo (ESA), e como nos gânglios da base (a) existe uma leptomeninge (L2) mais interna que impede o contacto direto do EVR com a artéria, ao contrário do que acontece no córtex (b). Adaptado de [22].

Os EVR encontram-se preenchidos de líquido intersticial e LCR, e atuam como drenos dos fluidos do sistema nervoso central para o sistema linfático, já que este não existe no encéfalo [2], [4].

O líquido dos EVR é, por sua vez, povoado por células da leptomeninge, mas também por macrófagos perivascularares, o que faz com que os EVR também desempenhem um papel ao nível da regulação imunológica e da barreira hematoencefálica. Os organismos invasores atravessam os vasos sanguíneos e chegam aos EVR, onde são destruídos pelos fagócitos. No entanto, também se estuda a possibilidade de doenças autoimunes do sistema nervoso, como a esclerose múltipla, estarem relacionadas com estes espaços, uma vez que pode ser este o local de ativação dos linfócitos T pelos macrófagos, levando ao processo inflamatório [21], [23].

2.3.1. Dilatação dos Espaços de Virchow-Robin

Tipicamente, os EVR são estruturas microscópicas, no entanto podem ficar dilatados, com dimensões na ordem dos milímetros (mm), o que os torna visíveis em imagens de RM [4].

Nestas condições, o diâmetro mede, geralmente, menos de 3 mm, mas pode medir até 15 mm em casos excepcionais, e o comprimento mede entre 3 e 15 mm [3], [24].

Apesar de a presente dissertação se debruçar apenas sobre os EVR dilatados com as dimensões supramencionadas, pode-se referir que também já foram observados casos extremamente dilatados, conhecidos como EVR gigantes, associados a efeitos de massas, que podem ser identificados, erroneamente, como neoplasias [25].

As causas que provocam dilatação dos EVR ainda não são completamente conhecidas, mas existem várias possibilidades [26]. Uma hipótese relaciona-se com o papel imunológico dos EVR durante um processo inflamatório, uma vez que a concentração de macrófagos perivasculares e linfócitos T aumenta muito nestes compartimentos [23]. Uma anomalia na permeabilidade das paredes das artérias, assim como um problema na drenagem do líquido intersticial, ou ainda a obstrução dos EVR também foram sugeridos [27]–[29]. O *stress* mecânico provocado por tensão arterial elevada é uma hipótese a ser considerada. Em idosos, a possibilidade da diminuição das dimensões do cérebro também foi avançado, uma vez que pode criar tuneis cheios de líquido intersticial ou LCR [27].

2.3.2. Identificação de Espaços de Virchow-Robin em imagens de ressonância magnética e localizações mais comuns

Como já foi referido, os EVR dilatados, tipicamente, surgem com diâmetros até 3 mm, e comprimento entre 3 mm e 15 mm. Relativamente à forma, podem ser identificados como pequenas estruturas tubulares, que quando observadas na direção do seu eixo apresentam formas arredondadas ou ovais, ou ainda, nos casos mais dilatados, com formas irregulares. Uma vez que os EVR rodeiam as artérias perforantes do cérebro, então a sua direção é perpendicular à superfície do parênquima cerebral [2], [8], [22], [30].

Os EVR podem ser identificados em imagens de RM, onde, visualmente, são isointensos relativamente ao LCR em todas as sequências de pulso. Porém, se se analisar os níveis de intensidades verifica-se que não são iguais, o que se deve ao facto de os EVR também conterem líquido intersticial, e por causa do efeito de volume parcial ser significativo, uma vez que se tratam de pequenas estruturas com dimensões próximas da resolução dos equipamentos de RM atuais [2], [4]. Assim sendo, o sinal destas estruturas em T1 e FLAIR é hipointenso, contrapondo-se ao sinal hiperintenso de T2. Adicionalmente, nas imagens de difusão não demonstram ter difusão restrita, pois são estruturas que comunicam ao longo dos vasos sanguíneos, nem se tornam mais intensas após injeção de contraste [2], [3], [26], [31].

Tipicamente, há locais onde a observação de EVR é mais comum, de tal forma que alguns autores as catalogaram em três tipos. No tipo I os espaços seguem a direção das *lenticulostriate arteries*, que atravessam a substância perfurada anterior e entram nos gânglios da base. O tipo II, por sua vez, surge associado às *perforating medullary arteries* que entram no córtex e se estendem até à SB. Por fim, o tipo III surge no mesencéfalo [2]. Quando os EVR se encontram predominantemente nos gânglios da base são designados por *état criblé* [32]. Na Figura 2.8 podem-se consultar alguns exemplos de EVR nos três tipos de localização referidos.

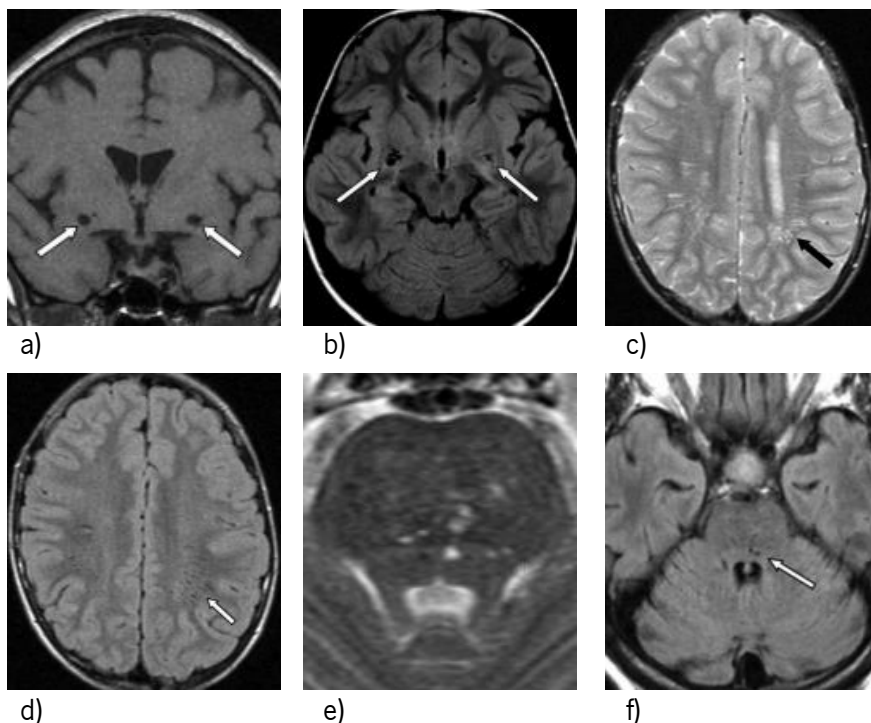


Figura 2.8 - Exemplos de Espaços de Virchow-Robin nos três tipos de locais mais comuns. Em a) e b) observam-se EVR nos locais do tipo I nas sequências T1 e FLAIR, respetivamente. Exemplos do tipo II correspondem às imagens c) e d), nas sequências T2 e FLAIR, respetivamente. e) é uma sequência T2 com ampliação e f) é uma sequência FLAIR, ambas exemplificando o tipo III. Apenas a) representa uma secção coronal do cérebro, enquanto as restantes são secções axiais. Imagens adaptadas [2].

Para além dos locais supramencionados, também podem ser identificados EVR no tálamo, cerebelo, córtex insular, região subinsular, hipocampo e SB em geral [22], [33].

2.3.3. Espaços de Virchow-Robin como biomarcadores de algumas doenças

A relevância clínica dos EVR dilatados é controversa, porque estes podem ser visíveis em cérebros saudáveis e em pessoas jovens, mas é mais comum serem observados em pessoas de idade avançada [22]. Adicionalmente, a técnica de RM empregue na aquisição das imagens também se revela importante, uma vez que a utilização de campos magnéticos fortes melhora a relação sinal-ruído, permitindo uma resolução espacial superior, melhorando a visualização e

identificação de EVR. Também a aquisição de volumes com mais *slices* (menor distância inter-*slice*) pode contribuir para o aumento do número de EVR detetados [2][4]. A aquisição de sequências de RM diferentes entre estudos dificulta a validação dos resultados entre eles. Por outro lado, a distinção entre EVR e lacunas pode ser difícil, o que pode causar erros na identificação dos mesmos [3].

No entanto, vários estudos correlacionam os EVR dilatados com doenças dos pequenos vasos do cérebro, particularmente com AVCs lacunares, doenças cognitivas e esclerose múltipla [2].

2.3.3.1. Doença dos pequenos vasos cerebrais

As doenças dos pequenos vasos cerebrais referem-se aos processos patológicos que atingem vasos sanguíneos de pequenas dimensões, como pequenas artérias e veias, arteríolas e capilares. Mais especificamente, a arteriosclerose diminui a elasticidade das paredes dos vasos sanguíneos, enquanto a angiopatia amiloide cerebral se caracteriza pela acumulação de amiloide nas paredes dos vasos, o que pode obstruir o fluxo sanguíneo e até mesmo levar à rutura dos vasos, por aumento da pressão a montante da obstrução [34].

Num estudo conduzido por *Rouhl et al*, em 2008, foram utilizadas imagens de RM de um grupo de 165 pessoas, que registaram o seu primeiro AVC lacunar [3]. Este tipo de AVC é derivado da oclusão de uma artéria perforante do cérebro, resultando em lesões de pequenas dimensões, com diâmetro entre 3 mm e 15 mm. Podem ser sintomáticos, traduzindo em manifestações físicas e/ou cognitivas evidentes, ou silenciosos, isto é, não tem consequências óbvias, podendo manifestar-se em alterações de humor e/ou cognitivas, mas as pessoas não têm consciência que os tiveram [32].

No estudo verifica-se que existe uma forte associação entre o número de EVR nos gânglios da base e a severidade das doenças dos pequenos vasos. Particularmente, esta associação existe relativamente aos AVCs lacunares silenciosos e lesões da SB. Porém, os autores referem limitações do estudo como a aquisição de sequências de RM diferentes de outros estudos, ou os critérios de distinção entre os EVR e os AVCs lacunares [3].

Em [30] verifica-se, novamente, que os EVR, totais e localizados nos gânglios da base, se correlacionam com os AVCs lacunares, mas não com os AVCs corticais. Também foi constatada a associação com hiperintensidades da SB profunda e periventricular. *Zhu et al* voltam a reforçar as referidas correlações, em particular nos gânglios da base, defendendo que

os EVR em RM devem ser considerados marcadores de doenças dos pequenos vasos, em idosos [35].

2.3.3.2. Problemas cognitivos

Os EVR dilatados foram associados com perdas de capacidades cognitivas. Ao nível das capacidades de raciocínio não-verbal e visuo-espaciais, verificou-se que existe uma relação inversa com o número de EVR dilatados. Apesar de a associação ter sido moderada, foi estatisticamente significativa, estando mesmo ao nível da associação dos mesmos problemas com lesões da SB. Porém, os autores admitem que o estudo pode não ser generalista, uma vez que apenas foram recrutados homens, de forma não aleatória. Porém, defendem que os EVR devem ser incluídos em futuros estudos cognitivos [36].

Foi também observado que a identificação de EVR dilatados nos gânglios da base pode ser útil na diferenciação entre demência vascular e demência degenerativa, como o Alzheimer. De facto, os autores utilizam uma medição semi-quantitativa do número de EVR, e verificaram que a sua presença no núcleo lentiforme (mais de 10) e no núcleo caudado permitem diferenciar as duas demências com sensibilidade de 67% e especificidade de 70%. Se se restringir o critério para mais de 10 no núcleo lentiforme e mais de 5 no núcleo caudado, a sensibilidade é reduzida para 25%, mas a especificidade sobe para 100%. Os autores concluem que os EVR dilatados têm relevância clínica, mas não se devem excluir outros marcadores, como a atrofia cerebral [4].

2.3.3.3. Outras relações

Para além da relação já descrita entre os EVR dilatados e as doenças dos pequenos vasos sanguíneos, também outras relações foram investigadas.

Assim, foi investigada a possível correlação entre a dilatação dos EVR e a esclerose múltipla. Verificou-se que o número de EVR visíveis em RM não permite diferenciar entre o grupo de pessoas com esclerose múltipla e o grupo de sujeitos normais. No entanto, o volume total dos EVR dilatados nos doentes com esclerose múltipla é significativamente maior do que no grupo saudável [21].

Num outro estudo foi sugerido que o número de EVR dilatados nos gânglios da base pode ser um indicador da resistência à terapia da depressão, em idosos. Foi apurado que, a observação dos EVR permite fazer o diagnóstico com 80% de sensibilidade e 62% de especificidade. No entanto são apontadas algumas limitações ao estudo, como a utilização de um sistema semi-quantitativo de contagem dos EVR desenvolvido para demência, e o número

reduzido de sujeitos de estudo (29 responderam bem ao tratamento e 21 desenvolveram resistência) [37].

A hipertensão arterial também parece poder estar relacionada com o aumento do número de EVR dilatados [35].

2.3.4. Métodos de contagem dos Espaços de Virchow-Robin dilatados

Atualmente, a investigação da correlação entre o número de EVR dilatados com outras doenças é feita com o recurso a contagens semi-quantitativas. Assim, as contagens são compostas por vários níveis, que correspondem a um *score* ou grau de severidade do número de EVR dilatados; cada nível corresponde a um intervalo de números de EVR dilatados observados, podendo, ainda, ser incluída informação do local.

A não existência de uma forma e critério de contagem *standard* é um dos motivos de discussão das conclusões, referentes à correlação dos EVR dilatados com doenças. Todos os autores encontrados dividiam a contagem por regiões do encéfalo, mas enquanto uns refinavam a contagem ao incluir informação de estruturas contidas na região em estudo [4], outros fazem a contagem analisando a região de maneira geral [3], [30], [35]. Na Tabela 2.3 e Tabela 2.4 pode-se consultar os critérios e formas de contar utilizados na bibliografia.

Tabela 2.3 - Métodos semi-quantitativos de contagem dos Espaços de Virchow-Robin dilatados. Os métodos diferem com a região do cérebro, e o critério de divisão de classes também varia com o estudo. A referência bibliográfica de cada escala é indicada na primeira coluna. SI – *substantia innominata*; NL – núcleo lentiforme; NC – núcleo caudado.

Ref.	Região	Critério/Nº de EVR	Score/severidade
		Nenhum	0
	Região subinsular	≤5 em ambos os lados	1
		>5 em um, ou ambos os lados	2
		Apenas na SI, e ≤5 em ambos os lados	0
	Gânglios da base – esquema 1	>5 na SI em ambos os lados, ou pelo menos 1 no NL	1
		Existência no núcleo caudado.	2
[4]		Apenas na SI, e ≤5 em ambos os lados	0
		Apenas na SI, e >5 em ambos os lados	1
	Gânglios da base – esquema 2	≤5 no NL em ambos os lados	2
		>5 e ≤10 no NL, ou ≤5 no NC, em ambos os lados	3
		>10 no NL e ≤5 no NC, em ambos os lados	4
		>10 no NL e >5 no NC, em ambos os lados	5

Tabela 2.4 - Métodos semi-quantitativos de contagem dos Espaços de Virchow-Robin dilatados (continuação). Os métodos diferem com a região do cérebro, e o critério de divisão de classes também varia com o estudo. A referência bibliográfica de cada escala é indicada na primeira coluna.

Ref.	Região	Critério/Nº de EVR	Score/severidade
[3]	Gânglios da base e SB cerebral	≤20	Baixo
		>20 e ≤50	Moderado
		>50	Alto
[35]	Gânglios da base	≤5	1
		>5 e ≤10	2
		>10, mas numerável	3
		Inumerável	4
[30]	Gânglios da base e SB cerebral	Nenhum	0
		≤10	1
		>10 e ≤20	2
		>20 e ≤40	3
		>40	4

2.4. Sumário

A RM assume-se como a modalidade imagiológica que melhor permite avaliar a estrutura dos tecidos moles do encéfalo, que é constituído por SC e SB. A SC encontra-se localizada na região mais superficial, designado córtex, mas também existem importantes núcleos de SC na região subcortical, os gânglios da base. Por seu lado, a SB preenche o restante espaço subcortical.

Como todos os tecidos, o encéfalo é irrigado por vasos sanguíneos, mas neste órgão estes são envolvidos por uma ou duas membranas, formadas a partir das meninges. O espaço entre a parede do vaso e as membranas designa-se por EVR. Por vezes, este fica dilatado, existindo vários estudos que correlacionam essa condição com várias doenças ou problemas de saúde, como os AVCs ou demência vascular, existindo a possibilidade de ser um novo biomarcador das mesmas. Os mesmos estudos focam a sua contagem em duas principais regiões de interesse, a SB, ou centro semioval, e os gânglios da base, utilizando escalas semi-quantitativas. A dificuldade de identificação destas estruturas, devido à sua reduzida dimensão, efeito de volume parcial e semelhança com alguns tipos de lesões, torna a sua contagem subjetiva e dependente do profissional que a realiza.

Capítulo 3 Estado da arte

A segmentação manual de imagens é um trabalho moroso, e como as aquisições de RM têm vários *slices* por sujeito torna-se ainda mais demorado. Além disso, é afetada pela subjetividade de quem a realiza, não tendo uma taxa de erro constante e conhecida. Assim, existem vários métodos de segmentação automática ou semiautomática, que tentam solucionar esses problemas.

Neste sentido, ao longo deste capítulo irão ser apresentados os trabalhos existentes, quer para a deteção automática de EVR, quer para a segmentação de tecidos do encéfalo.

A pesquisa de algoritmos de segmentação de tecidos do encéfalo tornou-se relevante quando se pretendeu implementar uma alternativa para a extração de SB, como região de interesse. Para isso, é necessário segmentá-la, devendo-se também segmentar a SC e LCR. Este é um problema com inúmeras soluções publicadas, portanto optou-se por seleccionar as principais linhas de investigação, dentro das quais se escolheu os trabalhos que se considerou mais relevantes.

Pelo contrário, para a deteção de EVR em imagens de RM irá apresentar-se todos os trabalhos possíveis, uma vez que este é o principal tema da presente dissertação.

3.1. Segmentação de tecidos do encéfalo

A segmentação dos tecidos do encéfalo em imagens de RM é um problema sobre o qual já se debruçaram diversos trabalhos, existindo algumas ferramentas disponíveis que o fazem. No entanto, é um problema que ainda se mantém atual, com publicações recentes sobre o mesmo e concursos internacionais.

Alguns autores abordaram o problema da segmentação do encéfalo utilizando o método de *clustering fuzzy c-means*, que consiste na minimização de uma função objetivo, atribuindo a cada *voxel* um grau de semelhança com cada classe, ao invés de classificar cada elemento com uma só classe. Posteriormente, no final das iterações, a classificação é feita atribuindo a cada *voxel* a classe com maior grau de semelhança, isto é, aquela classe cujas características do seu centróide se encontram mais próximas das características do *voxel*. Assim, em 1999, *Pham et*

al. utilizaram o *fuzzy c-means* adaptado para imagens 3D, para a utilização de mais do que uma sequência de RM e para corrigir a falta de homogeneidade. As principais vantagens apontadas pelos autores foram o algoritmo não ser supervisionado e ser robusto às condições iniciais, porém o facto de não incorporar informações da vizinhança foi uma desvantagem [38].

Quatro anos mais tarde, *Liew et al.* apresentaram um novo algoritmo baseado em *fuzzy c-means*, que para além de corrigir a falta de homogeneidade das imagens, também incorporava informação espacial através da vizinhança, utilizando um índice que media desigualdades entre os *voxels* e que foi adicionado à função objetivo. Assim, foi medida uma distância entre o *voxel* em análise e os seus vizinhos, que calculava o grau de semelhança entre eles, e caso fossem próximos então todos deviam ter um elevado grau de proximidade com a mesma classe. No entanto, caso a diferença entre o *voxel* central e a vizinhança fosse superior a um índice de diferença calculado para toda a imagem, o efeito da vizinhança era removido, pretendendo-se, assim, manter as bordas das regiões. Posteriormente, os autores testaram o método em imagens simuladas e reais, mas apenas existem resultados quantitativos para as primeiras, demonstrando sobreposições acima de 95% [39].

Já em 2009, *Wang et al.* alteraram o *fuzzy c-means* para conter informação da vizinhança, numa abordagem que utilizava várias escalas. A imagem era esbatida várias vezes com um filtro de difusão anisotrópico, e cada imagem obtida era representada numa nova escala. Em seguida, a segmentação começava pela imagem mais suavizada, utilizando o algoritmo *k-means*, e era propagada até à imagem original, que apresentava mais detalhes. Depois da inicialização com o *k-means* passava a utilizar-se o *fuzzy c-means* nas escalas seguintes, através da inclusão de um termo relacionado com a segmentação da escala anterior na função objetivo. Pode-se então dizer que nas escalas mais detalhadas o algoritmo era supervisionado pela segmentação anterior, o que o tornava robusto à presença de ruído, pois o seu efeito era minimizado nas imagens mais esbatidas. Além disso, também corrigia a falta de homogeneidade. Aplicando o método em imagens reais, adquiridas com um *scanner* de 4 T, obtiveram Dice de $85 \pm 3\%$, $82 \pm 5\%$ e $88 \pm 4\%$ para LCR, SC e SB, respetivamente [40].

As abordagens com modelos estatísticos são outra das classes de algoritmos utilizados na segmentação, sendo uma das alternativas ao *fuzzy c-means* apresentado anteriormente. Entre eles, o algoritmo *Expectation-Maximization* (EM) é muito popular para estimar e maximizar os parâmetros dos modelos, que depois permitem fazer a classificação [41]. Assim, enquanto as abordagens com *fuzzy c-means* utilizam a proximidade das características dos *voxels*, por

exemplo a intensidade, a um centróide da classe para a sua classificação, os algoritmos estatísticos têm um modelo para essas características, procurando-se aqueles *voxels* cujas características sejam melhor descritas pelo modelo de cada classe.

Em 1996, *Wells et al.* modelizaram a distribuição das intensidades de cada tecido através da distribuição gaussiana, mais precisamente uma mistura de gaussianas se forem considerados todos os tecidos, e de uma probabilidade *a priori* constante em todas as iterações. Além de estimarem a falta de homogeneidade com o EM, também sugeriram substituir a mistura de gaussianas por um método não paramétrico, na forma de uma janela de *Parzen*, que permitiria estimar a função densidade de probabilidades dos tecidos a partir de uma amostra de treino [42].

Van Leemput et al., em 1999, modelizaram as intensidades dos tecidos por uma mistura de gaussianas cujos parâmetros (média e variância) eram atualizados pelo EM, ao contrário de *Wells et al.*, enquanto a probabilidade *a priori* foi incluída através de *Markov Random Field* (MRF), utilizando o modelo de *Potts* para modelizar as relações de vizinhança. Desta forma, foi incorporado o contexto da vizinhança do *voxel*, ou seja, era uma probabilidade *a priori* que variava no espaço, o que ajudava a tornar a classificação mais coerente. Para além da aplicação do MRF, os autores ainda utilizaram um mapa de probabilidades para inicializar o algoritmo e restringir a probabilidade *a priori* em cada iteração. Os autores apontaram como principais desvantagens os *voxels* mal classificados devido ao efeito de volume parcial, ou a possibilidade de o MRF eliminar detalhes mais finos. Porém, testando em nove casos reais, os autores conseguiram mais de 94% de sobreposição em todos os casos, considerando todo o cérebro, e mais de 77% de sobreposição da SC [43]. Dois anos mais tarde, em 2001, *Zhang et al.*, propuseram um modelo baseado em modelos MRF, mas substituíram a mistura finita de gaussianas por um modelo *hidden MRF*. Também abandonaram a dependência de um mapa de probabilidades, apesar de ser possível incorporá-lo. Assim, os autores afirmaram que o seu algoritmo era mais robusto do que aqueles que se baseavam apenas em misturas de gaussianas, e que permitia adicionar a correção da falta de homogeneidade, em simultâneo com a classificação. Porém, foram apontadas como desvantagens o método de inicialização com *thresholds* e a possibilidade de o algoritmo EM encontrar um mínimo local, em vez de global; esta última desvantagem também se aplica aos trabalhos anteriores. Testando em nove RM de crianças, os autores obtiveram DICE entre 77,3% e 88,6% para SC em *slice* axial, e entre 74,4% e 86,4% para SC em *slice* coronal [44].

Em 2005, *Ashburner et al.* propuseram um método probabilístico intercalado com registo. Assim, faziam a segmentação dos tecidos e registo de um mapa de probabilidades, enquanto corrigiam a falta de homogeneidade. A distribuição das intensidades dos tecidos foi modelizada através de uma mistura de gaussianas, cujos parâmetros eram estimados com o EM, enquanto o *prior* era proveniente de mapas de probabilidades dos tecidos. Este último podia ser deformado, ao contrário das aplicações mais clássicas, porém não incorporava informações da vizinhança e os mapas eram dependentes da utilização pretendida, por exemplo não podiam ser utilizados mapas de adultos se o método fosse aplicado em crianças. Os autores testaram em imagens simuladas, e os melhores resultados foram obtidos com a sequência T1, para 0% de falta homogeneidade, com resultados em DICE como 0.933, 0.961 para SC e SB [45].

Alguns autores prescindiram da introdução da informação espacial através da probabilidade *a priori*, preferindo incorporá-la no modelo de misturas de gaussianas. Assim, *Greenspan et al.*, em 2006, utilizaram uma mistura com quatro dimensões, considerando a intensidade e as coordenadas da imagem (x,y,z). Para além disso, cada tecido compreendia muitas gaussianas, ao contrário dos métodos anteriores que só consideravam uma por tecido, de forma a codificar as várias localizações 3D. A classificação utilizava a regra de *Bayes*, depois de o EM estimar os parâmetros do modelo. Testando na base de dados *Internet Brain Segmentation Repository* (IBSR), os autores obtiveram os resultados, em Dice, de 0,81 e 0,795 para a SC e SB, respetivamente, afirmando estar ao nível da implementação de *van Leemput*, e sem utilizar MRFs [46].

Mais recentemente, o aprimoramento da segmentação da SC do córtex tornou-se alvo de investigação nas abordagens estatísticas, tal como foi apresentado no trabalho desenvolvido por *Cardoso et al.*, em 2011. Utilizando como base o trabalho de *van Leemput et al.*, já referido, os autores introduziram três alterações. Em primeiro lugar, depois de o algoritmo EM convergir, os mapas de probabilidades que servem de *priors* eram modificados para se adaptarem à variabilidade inerente a cada paciente, combinando as probabilidades *a posteriori* com a informação da classificação da vizinhança. Depois, corrigiam o efeito de volume parcial ao considerar duas novas classes (SC/SB e SC/LCR) que misturavam as gaussianas dos tecidos puros; criando também um novo mapa de probabilidades, tendo por base o modelo de tecidos puros. Por fim, utilizando as estimativas da probabilidade *a posteriori* era criado um mapa do sujeito que permitia identificar os sulcos e os giros do córtex, onde os *priors* e a força do MRF eram alterados, pois estes últimos são responsáveis por eliminar estruturas mais detalhadas.

Finalmente, os autores compararam os seu resultados com as técnicas descritas em [44], [45], [47], conseguindo sempre melhores resultados, relativamente ao córtex [41].

Song et al., em 2006, aplicaram a teoria de grafos à segmentação de tecidos cerebrais, com o método *Graph Cut*, ao considerar que os *voxels* da imagem estariam conectados aos seus vizinhos, e simultaneamente à classe à qual pertenciam. Assim, era possível minimizar uma função de energia proveniente de um MRF, que tinha em consideração o contexto da vizinhança e a observação dos dados. Para além disso, foi adicionado um peso às arestas que ligam os *voxels* às classes, proveniente de um mapa de probabilidades de tecidos, e um peso às arestas que liga os *voxels* vizinhos tendo em conta as intensidades e um mapa probabilístico de contornos, calculado com o gradiente da imagem. Finalmente, a segmentação era obtida quando o algoritmo *Graph Cut* minimizava a função de energia. Testando em imagens de recém-nascidos, que são mais difíceis de segmentar do que imagens de adultos, o autores obtiveram 0,745 e 0,769 para SC e SB, respetivamente, em Dice [48].

Alguns autores optaram pelo método *mean-shift*, por ser um método não paramétrico, ao contrário dos estatísticos, e não necessitar de saber o número de *clusters* para ser aplicado, contrapondo-se aos estatísticos e *fuzzy c-means*. Este algoritmo percorre cada ponto, e a partir daí cria um vetor que aponta para o local onde a densidade é maior, ou seja, onde existe uma moda, e agrupa no mesmo *cluster* os pontos que tendem para a mesma região [49], [50]. Os próximos dois trabalhos tem como base esta abordagem.

Na sua abordagem, *Jiménez-Alaniz et al.*, em 2006, começaram por calcular um mapa de confiança das bordas com o gradiente da imagem e suas orientações, para evitar que os *voxels* das fronteiras de tecidos fossem tomados como modas, durante a execução do *mean-shift*. Após a sua aplicação, de forma a diminuir o número de *clusters* detetados, agrupavam aqueles que se encontravam mais próximos e cuja fronteira era mais fraca. Assim, os mais pequenos eram agrupados aos maiores que se encontravam próximos. Por fim, selecionavam-se quatro sementes, entre os *clusters* (fundo, LCR, SC e SB), e com um classificador Bayesiano que utilizava como *prior* os mapas de probabilidades de tecidos, classificavam-se os restantes *clusters* tendo em conta a distância entre as suas intensidades e as das sementes, obtendo-se a segmentação final. As principais vantagens eram a robustez à presença de falta de homogeneidade na imagem e a capacidade para preservar fronteiras, mas foi adiantado que a dimensão da janela da vizinhança do *mean-shift* deveria ser escolhida com cuidado. Finalmente,

testando na base de dados do IBSR obtiveram os resultados, em Dice, de 0,347, 0,745 e 0,771 para o LCR, SC e SB, respetivamente [50].

Em 2009, *Mayer et al.*, aplicaram este método sobre as características da imagem, isto é, cada ponto foi definido pela intensidade em cada sequência de RM e localização 3D. Além disso, como a dimensão da janela de vizinhança utilizada na pesquisa da moda era variável, diz-se que foi utilizado um algoritmo *mean-shift* adaptativo. Posteriormente, e por existirem muitas mais classes do que as pretendidas, os autores agrupavam aquelas cujos centros se encontravam mais próximos. No terceiro, e último passo, os *clusters* resultantes da etapa anterior eram agrupados em três classes (SC, SB e LCR) através da utilização do algoritmo *k-means*. Já no que se refere a resultados, o método foi testado na base de dados do IBSR e obteve classificações, em Dice, de 0,811 e 0,817 para SC e SB, respetivamente [49].

Mais recentemente, a escolha de métodos supervisionados, que impliquem um passo de aprendizagem, tem começado a ganhar relevo, com autores a optarem por abordagens com *K-Nearest Neighbor*, *Support Vector Machines* ou *Random Decision Forests*, que serão apresentados de seguida.

Anbeek et al., em 2005, utilizaram o método *K-Nearest Neighbor* que consiste em procurar no conjunto de supervisão os pontos cujas características estão mais próximas do ponto em estudo, e classificá-lo de acordo com a classe da maioria. Porém, no trabalho em questão, os autores não classificavam definitivamente os *voxels*, antes criavam mapas de probabilidades para cada tecido, dividindo, em cada *voxel*, o número de K vizinhos de cada classe pelo conjunto de todos os K vizinhos considerados, e posteriormente aplicavam um *threshold* sobre os mapas. As características que os autores consideraram foram as intensidades em todas as sequências de pulso e as posições dos *voxels* em cada eixo (x,y,z), concluindo que estas últimas eram relevantes para separar os ventrículos do restante LCR [51].

Akselrod-Ballin et al., em 2006, optaram por *Support Vector Machines*. No entanto, antes da fase de classificação, os *voxels* eram agregados em conjuntos, numa abordagem a várias escalas. Assim, cada *voxel* era agregado com os que se encontravam adjacentes, com base na semelhança das intensidades, sendo atribuído um peso; em seguida os agregados já definidos estabeleciam ligações com os mais semelhantes, e com um novo peso associado, e assim sucessivamente, numa estrutura piramidal. A formação dos agregados tinha por base um conjunto de características calculadas com os *voxels* que lhes pertenciam, como a média das intensidades, a intensidade máxima, variância, momentos das intensidades, localização (x,y,z),

forma, área da superfície, média das intensidades dos agregados vizinhos e mapa de probabilidades de tecidos. Depois, as características das escalas 5 e 6 eram introduzidas no classificador, os aglomerados das escalas eram classificados em separado e o resultado era projetado nos *voxels* através de matrizes de interpolação, tendo em conta os pesos das ligações. Testando o algoritmo nas imagens do IBSR, os autores obtiveram resultados, em DICE, de 0,514, 0,81 e 0,802 para o LCR, SC e SB, respetivamente, demonstrando ter um método competitivo [52], [53].

Por seu lado, *Yi et al.*, em 2009, abordou o problema utilizando outro classificador supervisionado, as *Random Decision Forests*. Após a correção da falta de homogeneidade da imagem, eram extraídas as características que seriam usadas na classificação. O autor considerou que seria relevante o classificador tomar em consideração as intensidades da imagem, o módulo do gradiente, as probabilidades provenientes de um atlas, a probabilidade de verosimilhança de uma mistura de gaussianas treinada e as probabilidades *a posteriori* de cada classe. Depois de classificar, eram corrigidos os *voxels* que sofriam do efeito de volume parcial. Os autores assumiram como principais vantagens a capacidade de generalização das *Random Decision Forests*, e a incorporação de informação espacial através do atlas, em contrapartida o classificador não tinha, diretamente, em conta os *voxels* da vizinhança. Utilizando apenas sequências T1 da base de dados IBSR, foram atingidos resultados DICE de 0.699, 0.900 e 0.831 para o LCR, SC e SB, respetivamente [9].

3.2. Segmentação e identificação de Espaços de Virchow-Robin dilatados

Na literatura, são poucos os trabalhos que utilizaram métodos computadorizados para a identificação de EVR dilatados, e desses, alguns detetaram-nos como falsos positivos na segmentação de pequenas hiperintensidades em T2, sendo depois removidos.

Em 2004, *Descombes et al.* aplicaram um método para a deteção e contagem de EVR dilatados, exclusivamente. Começavam por filtrar a imagem de forma a realçar alguns aspetos dessas estruturas, como a largura permitida e o contraste, de maneira a que a pesquisa de candidatos a EVR fosse facilitada. Em seguida, era utilizado um *Marked Point Process* para associar o comprimento e orientação a cada candidato. Além disso, era possível incorporar informação geométrica e de localização, assim como informação proveniente dos dados (imagem). O modelo definido pelo *Marked Point Process* era, então, otimizado através da utilização do algoritmo *Reversible Jump Markov Chain Monte Carlo*. Aplicando apenas à

sequência T1, cuja aquisição foi 3D, sendo, posteriormente, interpolada para uma resolução isotrópica de 1 mm, o método identificou 668 EVR dilatados, em 7 volumes, contra 597 marcados pelos médicos [24].

Quatro anos mais tarde, em 2008, *Wuerfel et al.* optaram por uma abordagem semiautomática, baseada em *thresholds* e na conectividade das regiões, para identificar os EVR dilatados, que seriam alvo de contagem e medição do seu volume. Foram utilizadas as sequências T1 3D, T2 e FLAIR, estando co-registadas. O foco dos autores era o estudo da correlação entre os EVR dilatados e a esclerose múltipla. A avaliação do método semiautomático demonstrou que a correlação entre o número de EVR dilatados detectados e o número reportado pelo observador, em 10 casos aleatórios, foi de 0,996, que subiu para 0,998 quando se mediu o volume dos mesmos [21].

Por seu lado, *Uchiyama et al.*, no mesmo ano, utilizaram uma operação morfológica *white top-hat* na sequência T2 para enfatizar pequenas regiões hiperintensas, que eram depois segmentadas aplicando um *threshold*. Os autores apenas estavam interessados na extração de lacunas, mas sendo estas e os EVR dilatados estruturas de pequenas dimensões, ambos eram extraídos. Para fazer a distinção, os autores utilizaram uma rede neuronal, cujas características eram a localização, dimensão, irregularidade e diferença entre a intensidade em T1 e T2 [54].

Num outro trabalho, publicado em 2011, *Ramirez et al.* segmentaram hiperintensidades subcorticais, incluindo os EVR dilatados, utilizando um *threshold* adaptativo. Os EVR eram, mais tarde, excluídos através da remoção de componentes ligados de pequenas dimensões, ou manualmente. Foram utilizadas as sequências T1, T2 e *Proton Density* (PD) [55].

3.3. Sumário

A segmentação dos tecidos do encéfalo é uma tarefa que já é foco de estudo há mais de 15 anos, existindo diversas abordagens. Predominantemente, utilizaram-se algoritmos não-supervisionados, com especial destaque para os estatísticos. No entanto, nos últimos anos, já têm sido estudadas abordagens que utilizaram classificadores supervisionados, com resultados competitivos. Na Tabela 3.1 pode-se consultar o resumo do estado da arte. Comparar este tipo de algoritmos não é trivial, porque a maioria utiliza imagens de RM privadas, e as métricas também variam de trabalho para trabalho. Porém, alguns autores optam por utilizar a base de dados pública IBSR e publicar os seus resultados com as métricas Dice ou Jaccard, razão pela qual só se incluíram estes resultados na Tabela 3.1.

Por outro lado, o problema da detecção computadorizada de EVR foi pouco estudado, tendo-se encontrado somente quatro trabalhos publicados, e metade apenas os deteta como erros. Destes, o trabalho de *Descombes et al, 2004*, é o único totalmente automático. Na Tabela 3.2 encontra-se o resumo do estado da arte para estes algoritmos.

Tabela 3.1 - Estado da arte de segmentação de tecidos. Na coluna Abordagem encontra-se o grupo geral de cada algoritmo, na coluna Particularidade especifica-se o fator inovador, em Observações fornecem-se mais informações relevantes e na coluna Resultados encontram-se os resultados, em DICE, apenas para os algoritmos que utilizaram a base de dados IBSR.

Abordagem	Trabalho	Particularidade	Observações	Resultados (Dice)
Fuzzy c-means	<i>Pham et al, 1999</i> [38]	-	Robusto às condições iniciais. Não tem em conta a vizinhança.	-
	<i>Liew et al, 2003</i> [39]	Introduz informação da vizinhança.	Estratégia para preservar bordas.	-
	<i>Wang et al, 2009</i> [40]	Abordagem multi-escala.	Utiliza a vizinhança. Robusto à presença de ruído.	-
Modelos estatísticos	<i>Wells et al, 1996</i> [42]	Mistura de gaussianas.	Probabilidade <i>a priori</i> estacionária.	-
	<i>Van Leemput et al, 1999</i> [43]	Mistura de gaussianas e MRF.	Utilização de atlas de probabilidades. MRF pode eliminar detalhes finos.	-
	<i>Zhang et al, 2001</i> [56]	<i>Hidden MRF</i> .	Abandona o atlas de probabilidades, mas ressalva que pode ser incorporado	-
	<i>Ashburner et al, 2005</i> [45]	Intercala classificação e registo de atlas, iterativamente.	Mistura de gaussianas como modelo probabilístico. Probabilidade <i>a priori</i> proveniente de atlas.	-
	<i>Greenspan et al, 2006</i> [46]	Informação espacial incorporada na mistura de gaussianas.	-	SC: 0,81 SB: 0,795
	<i>Cardoso et al, 2011</i> [41]	Abordagem de [43] + adaptação do atlas ao sujeito.	Dedicado à segmentação do córtex. Melhoria nos sulcos e giros.	-
	Grafos	<i>Song et al, 2006</i> [48]	<i>Graph cuts</i> .	Função de energia de MRF.
Mean-shift	<i>Jimenez-Alaniz et al, 2006</i> [50]	<i>Mean-shift</i> + classificador Bayesiano.	Utiliza atlas para probabilidades <i>a priori</i> .	LCR: 0,347 SC: 0,745 SB: 0,771
	<i>Mayer et al, 2009</i> [49]	Localização 3D como característica.	Janela de procura adaptativa.	SC: 0,811 SB: 0,817
Supervisionada	<i>Anbeek et al, 2005</i> [51]	<i>K-nearest neighbor</i> .	Classificador cria mapas de probabilidades.	-
	<i>Akselrod-Ballin, 2006</i> [52], [53]	<i>Support Vector Machines</i> .	Os <i>voxels</i> são agregados antes da classificação.	LCR: 0,514 SC: 0,81 SB: 0,802
	<i>Yi et al, 2009</i> [9]	<i>Random Decision Forests</i> .	-	LCR: 0,699 SC: 0,9 SB: 0,831

Tabela 3.2 - Estado da arte de algoritmos de detecção de Espaços de Virchow-Robin.

Trabalho	Exclusividade para EVR	Abordagem
<i>Descombes et al, 2004</i> [24]	Sim	Automática. Modelização da distribuição e interação dos EVR com <i>Marked Point Process</i> e otimização com RJMCMC
<i>Wuerfel et al, 2008</i> [21]	Sim	Semi-automático. Baseado em <i>thresholds</i> e conectividade de regiões
<i>Uchiyama et al, 2008</i> [54]	Não	<i>White top-hat</i> e <i>thresholds</i> . Separação de EVR e lacunas com rede neuronal.
<i>Ramirez et al, 2011</i> [55]	Não	<i>Threshold</i> adaptativo.

Capítulo 4 Modelização através de *Point Process*

Um Processo Estocástico, ou aleatório, é um conjunto de variáveis aleatórias indexadas, que permite estudar a sua evolução, geralmente ao longo do tempo. Um *Point Process* (PP) é um processo estocástico que descreve a dispersão de um agrupamento de pontos, num dado espaço, ou seja, cada variável aleatória corresponde a uma realização de um conjunto de pontos, em locais aleatórios. É utilizado para estudo de padrões, tipicamente, em 1, 2 ou 3 dimensões [57], [58].

Os pontos representam as localizações dos acontecimentos, e a análise estatística deste tipo de padrões visa descrever e modelizar as interações entre eles, assim como a sua localização. Também é possível adicionar uma marca aos pontos, correspondente a alguma propriedade dos objetos, o que permite estudar ainda melhor os processos que estão a gerar os padrões, designando-se, neste caso, como *Marked Point Process* (MPP) [57], [58].

Assim, é possível criar modelos que descrevam um determinado problema. Neste contexto, os EVR podem ser caracterizados por um ponto, e respetiva marca, por exemplo, um vetor de comprimento que também informe acerca da sua direção. Pode-se, ainda, incluir no modelo termos de interação e *priors* de comprimento e adequação aos dados da imagem.

Neste capítulo, introduzir-se-á os conceitos de PP, começando pelo mais simples destes processos, o Poisson PP, uma vez que serve de base a muitos modelos. Em seguida, demonstra-se como se pode adicionar termos de interação entre pontos, através dos Gibbs PP. Avança-se, então, para os conceitos de MPP, dando-se o exemplo de um Gibbs MPP, utilizado para a extração de segmentos de reta numa imagem, sendo semelhante ao caso dos EVR.

4.1. Point Process

Um PP é um padrão de pontos numa unidade de medida, por exemplo área, numerados para efeito de notação, mas que não implica nenhuma ordem específica. Considerando que N representa um PP, então N é um conjunto de pontos, $N = \{x_1, x_2, \dots\}$, onde x_i é o ponto i . Também se utiliza a mesma nomenclatura para escrever a contagem de pontos de um subconjunto, por exemplo $N(B)$ é a contagem de pontos do subconjunto B , e se este for

limitado, então $N(B) < \infty$ [57], [58]. Na Figura 4.1 encontra-se um exemplo de PP, com um subconjunto marcado.

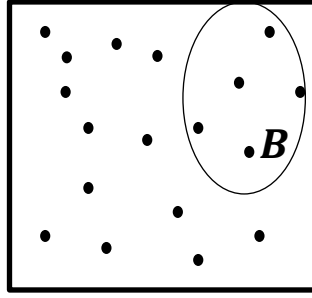


Figura 4.1 - Exemplo de *Point Process*. Neste caso, $N(B) = 5$.

Formalmente, se o espaço onde se fazem as medidas for definido por (K, \mathfrak{B}, ν) , onde K representa o subconjunto de \mathbb{R}^d onde se dá o PP, \mathfrak{B} é o conjunto de todos os subconjuntos de K e ν é uma medida do espaço d -dimensional de K , ou seja, é a medida do espaço, por exemplo, se $d = 3$ então a medida é volume. Considerando, ainda, o espaço de configurações $X = \bigcup_{n=0}^{\infty} X_n$, com σ -álgebra \mathcal{F} gerada pelos mapeamentos $x_1, \dots, x_n \rightarrow \sum_{i=1}^n \mathbf{1}\{x_i \in B\}$ que faz a contagem de pontos de conjuntos $B \in \mathfrak{B}$. Então, um PP em K é um mapeamento de um espaço de probabilidades para (X, \mathcal{F}) [59].

A medida de intensidade faz sentido em conjuntos fechados; neste caso existe a esperança que no conjunto B exista um número de pontos igual à medida de intensidade, ou seja, $\Lambda(B) = \mathbf{E}(N(B))$. Assim, pode-se definir uma função densidade $\lambda(x)$, sendo a função de intensidade do processo a Equação (4.1) [57], [58].

$$\Lambda(B) = \int_B \lambda(x) dx \tag{4.1}$$

Quando se quer escrever a probabilidade de um dado evento, por exemplo, a probabilidade de existirem n pontos no conjunto B , faz-se $P(N(B) = n)$ [57].

Existem PP finitos e infinitos. O primeiro caso acontece quando existe uma barreira clara que limita o conjunto dos pontos, já no segundo essa barreira não existe, ou não é clara. No presente trabalho está-se perante um PP finito, limitado pelo limite das regiões de interesse [57].

Se for possível fazer a translação de um PP, e mesmo assim ele obedecer à mesma distribuição, então é estacionário (Equação (4.2)).

$$\begin{aligned} P(N(B_1) = n_1, \dots, N(B_k) = n_k) \\ = P(N(B_1 - x) = n_1, \dots, N(B_k - x) = n_k) \end{aligned} \tag{4.2}$$

Desta forma a intensidade fica simplificada, como se observa na Equação (4.3), uma vez que λ não é dependente do local onde se encontra, sendo igual em todo o espaço, ficando a intensidade apenas dependente da proporção de espaço ocupado pelo conjunto de pontos [57], [58].

$$\Lambda(B) = \lambda\nu(B) \quad (4.3)$$

4.1.1. Poisson Point Process

Os Poisson PPs estacionários são adequados quando os pontos se distribuem de forma totalmente aleatória. Para além disso, servem de base à construção de modelos mais complexos. Nestes casos, o número de pontos num conjunto fechado B segue uma distribuição de Poisson com intensidade $\lambda\nu(B)$, sendo λ constante, e o número de pontos em conjuntos disjuntos formam variáveis aleatórias independentes [57].

Assim sendo, considerando as duas propriedades supramencionadas, a distribuição de Poisson de um PP, num conjunto fechado, é descrito pela Equação (4.4).

$$P(N(B) = n) = \frac{\lambda^n (\nu(B))^n}{n!} e^{-\lambda\nu(B)} \quad (4.4)$$

É comum ver-se a distribuição aplicada a uma dimensão, assim $\nu(B)$ representaria um intervalo de tempo, mas também pode ser usada para várias dimensões, por exemplo se for bidimensional então $\nu(B)$ representa a área que delimita o conjunto B . Se, por outro lado, existirem k conjuntos disjuntos, então a distribuição conjunta é a da Equação (4.5) [57].

$$\begin{aligned} P(N(B_1) = n_1, \dots, N(B_k) = n_k) \\ = \frac{\lambda^{n_1 + \dots + n_k} (\nu(B_1))^{n_1} \dots (\nu(B_k))^{n_k}}{n_1! \dots n_k!} e^{-\sum_{i=1}^k \lambda\nu(B_i)} \end{aligned} \quad (4.5)$$

Uma importante condição dos Poisson PPs é a sua dispersão independente, isto é, não existe interação entre os pontos do padrão [57].

4.1.2. Gibbs Point Process

Os Gibbs PPs são PPs finitos, que tomam em consideração a interação entre pares de pontos [57].

Considerando n pontos fixos numa janela W , com $x_1, \dots, x_n \in W$, a função densidade de probabilidades das localizações, não dependentes da ordem dos pontos, é descrita pela Equação (4.6) [57].

$$f_n(x_1, \dots, x_n) = \frac{1}{Z_n} e^{-\sum_{i=1}^{n-1} \sum_{j=i+1}^n \phi(\|x_i - x_j\|)} \quad (4.6)$$

onde $\phi(r)$ é a função de potencial da interação entre pares, calculada com base na sua distância (r), e Z_n é a função de partição da configuração, que faz a normalização. O expoente correspondente à soma dupla é designado por energia total do sistema ($U(x_1, \dots, x_n)$) [57], [58].

Também é comum definir a interação entre os pontos através de uma função de interação (Equação (4.7)), o que faz com que a função densidade de localização se transforme na Equação (4.8) [57].

$$h(r) = e^{-\phi(r)} \quad (4.7)$$

$$f_n(x_1, \dots, x_n) = \frac{\prod_{i=1}^{n-1} \prod_{j=i+1}^n h(\|x_i - x_j\|)}{Z_n} \quad (4.8)$$

Caso os pontos não interajam, então $h(r) = 1$, e fica-se perante um PP de Poisson. Se, por outro lado, $h(r) < 1$ então a relação é de repulsão; no caso contrário existe atração e formação de aglomerados [57].

No caso de se considerar que o número de pontos e a sua localização é aleatório, então a densidade deve considerar o processo de Poisson (Π) com intensidade unitária, que os gere. Assim, considerando que $\mathbf{x} = \{x_1, \dots, x_n\}$ representa uma configuração de pontos, definido no espaço limitado N_{fin} , então a distribuição de um PP A é calculada pela Equação(4.9) [57].

$$P(N \in A) = \int_A p(\mathbf{x}) \Pi(d\mathbf{x}) \quad (4.9)$$

A necessidade de se considerar um processo de Poisson está relacionado com o facto de o número de pontos ser aleatório, ou seja, imagine-se que existe um PP totalmente aleatório e sem interação entre os pontos. Posteriormente, vai-se calcular a probabilidade desse conjunto de pontos seguir um PP de Gibbs, através da análise da interação entre eles. Por isso, a probabilidade é calculada através do integral ao longo do PP A , que é um subconjunto do PP de Poisson. Assim, se a probabilidade fosse calculada ao longo do conjunto do processo de Poisson seria $P(N \in N_{fin}) = \int_{N_{fin}} p(\mathbf{x}) \Pi(d\mathbf{x}) = 1$. A intensidade unitária do processo de Poisson é assumida, uma vez que se obteria o mesmo se fosse diferente, apenas a função de normalização (Z) da Equação (4.10) seria alterada [57].

O processo de Gibbs para um número aleatório de pontos, relativamente a um processo de Poisson de intensidade unitária, é definido pela densidade da Equação (4.10) [57].

$$p(\mathbf{x}) = \frac{1}{Z} e^{-(\alpha n + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \phi(\|x_i - x_j\|))} \quad (4.10)$$

onde Z é a função de normalização e α é o termo de atividade química.

4.2. Marked Point Process

No MPP, os pontos deixam de representar apenas uma localização no espaço, passando a ter associado uma marca, que se trata de uma característica do objeto localizado pelo ponto. Por exemplo, se se estiver a estudar a dispersão de árvores numa floresta pode ser conveniente associar, à localização de cada árvore, o raio da mesma. As marcas podem ser quantitativas, com valor real, (Figura 4.2(b)), ou qualitativas (Figura 4.2(a)). Além disso, o conjunto das marcas pode ser finito ou infinito, ou ainda discreto, se for qualitativa, ou contínuo, no caso de ser quantitativa [57], [58].

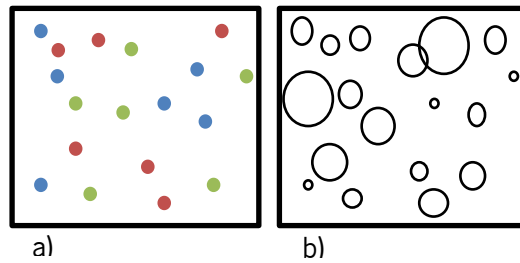


Figura 4.2 – *Marked Point Process* com marcas qualitativas (cor) e marcas quantitativas (raio). Em a) o conjunto das marcas é qualitativo (cor), discreto e finito (azul, vermelho e verde) enquanto em b) o raio é quantitativo, podendo ter qualquer medida, pelo que é contínuo, podendo ser infinito.

Portanto, sendo N um PP no subconjunto $B \subseteq \mathbb{R}^d$, de tal forma que qualquer ponto $x_i \in B$, se a cada ponto estiver associada uma marca aleatória, $m(x_i) \in C$, com $C \subseteq \mathbb{R}$, então está-se perante um MPP $M = \{[x_i; m(x_i)]\} = \{[x_1; m(x_1)], [x_2; m(x_2)], \dots\}$ definido em $B \times C$. Neste caso, a intensidade do processo pode ser calculada utilizando a Equação (4.11) [57].

$$\Lambda(M(B \times C)) = \int_B \int_C \lambda(x, m) dm dx \quad (4.11)$$

À semelhança do que acontece com os PP, os MPP também podem ser estacionários, o que é verdade se existir uma translação dos pontos de \mathbf{x} , mantendo as marcas, de acordo com a Equação (4.12) [57].

$$M = \{[x_i, m(x_i)]\} \triangleq M_x = \{[x_i + x, m(x_i)]\}, \text{ com } x \in \mathbb{R}^d \quad (4.12)$$

O próprio modelo de Gibbs pode ser adaptado para tirar partido das marcas dos pontos, ao modificar a função de potencial para considerar as marcas nas relações entre os pontos, ficando $\phi(m_1, m_2, r)$ [57].

4.2.1. Gibbs Point Process adaptado para a extração de segmentos de reta

Stoica et al. extraíram estradas a partir de imagens de satélite através da modelização de segmentos de reta, utilizando um Gibbs *Point Process*. Este incorporava interações entre os pontos, utilizando não só as suas localizações mas também as suas marcas, demonstrando que este tipo de modelos é versátil e podem ser adaptados para aplicações específicas [59], [60].

Partindo da forma geral do PP de Gibbs, considera-se, neste caso, que a sua energia total deve-se a dois tipos, a energia de interação, U_I , e a energia dos dados, U_D , tal como se encontra na Equação (4.13), sendo \mathbf{s} o conjunto dos pontos e das suas marcas de tal forma que $\mathbf{s} = \{[x_1, m(x_1)], [x_2, m(x_2)], \dots\}$ [59], [60].

$$f(\mathbf{s}) \propto \beta^n e^{-U(\mathbf{s})} = \beta^n e^{-(U_D(\mathbf{s})+U_I(\mathbf{s}))} \quad (4.13)$$

onde β representa a intensidade de um processo de Poisson, não necessariamente unitário.

A interação entre os objetos segue o Modelo de *Candy*, cuja densidade, não normalizada, é apresentada na Equação (4.14) [59], [60].

$$f_I(\mathbf{s}) = \beta^n \prod_{s_i \in \mathbf{s}} g(s_i) \prod_{s_i \diamond s_j, i < j} h(s_i, s_j) \quad (4.14)$$

onde $s_i \diamond s_j$ significa que os segmentos s_i e s_j interagem, sendo essa interação modelizada pela função h , por exemplo, se dois segmentos estiverem próximos e alinhados, então existe energia de atração entre eles, mas também pode ser utilizado para outro tipo de interação. g é a função que modeliza o comprimento dos segmentos, ao considerar a marca comprimento associada a cada segmento de reta candidato. Uma vez que ambas as funções estão na forma e^U , aplicando o logaritmo natural ao Modelo de *Candy*, obtém-se a energia de interação (Equação (4.15)) [59], [60].

$$U_I(\mathbf{s}) = - \sum_{s_i \in \mathbf{s}} \left[\ln(g(s_i)) + \sum_{s_i \diamond s_j, i < j} \ln(h(s_i, s_j)) \right] \quad (4.15)$$

A energia proveniente dos dados verifica se um segmento pertence à rede ou não, através da análise das estatísticas do mesmo, extraídas com base na imagem, ou seja, nos *pixels* que o constituem, ou da vizinhança [59], [60].

Portanto, a configuração de pontos que minimizar a energia do modelo (Equação (4.16)) é aquela que, provavelmente, corresponde melhor à rede de segmentos que formam as estradas [59], [60].

$$\begin{aligned}
 U(\mathbf{s}) &= U_D(\mathbf{s}) + U_I(\mathbf{s}) \\
 &= - \sum_{s_i \in \mathbf{s}} \left[U_D(s_i) + \ln(g(s_i)) + \sum_{s_i \circ s_j, l < j} \ln(h(s_i, s_j)) \right] \quad (4.16)
 \end{aligned}$$

Descombes et al., em 2004, utilizaram uma abordagem semelhante para extrair os EVR de imagens de RM, ao modelizá-los como segmentos de reta.

4.3. Sumário

Um PP é uma realização de pontos numa dada medida do espaço, por exemplo área (2D) ou volume (3D), cada ponto correspondendo à sua localização. Porém, pode-se atribuir mais significado a cada elemento, associando-lhe uma marca. Cada realização está sujeita a uma distribuição, que a rege.

O Poisson PP é o mais simples dos processos, cuja distribuição subjacente é a distribuição de Poisson. Neste caso, não existe qualquer tipo de interações entre os pontos. No entanto, se numa dessas realizações existir interação, pode-se estar perante um PP de Gibbs. A vantagem deste PP é a possibilidade de se introduzir, na sua função de energia, vários termos, tendo-se dado destaque ao modelo de *Candy*, que foi introduzido com o propósito de integrar interações entre segmentos de reta. Este é importante, uma vez que o modelo utilizado para os EVR considera essas estruturas como pequenos segmentos, com orientação, modelando as suas propriedades de intensidade e interação através de um MPP.

A simulação é uma ferramenta importante no estudo e na aplicação de PP, uma vez que, com frequência, os cálculos são complexos, ou mesmo intratáveis. Assim, simulam-se padrões de pontos aleatórios, que aproximem o modelo que se pretende estudar, o qual deve ser conhecido ou assumido. Apesar de os resultados serem aproximados, podem, mesmo assim, ser muito precisos, geralmente à custa de carga computacional [57], [58].

Assim sendo, pode-se simular PPs para calcular características dos mesmos, para estudar o comportamento dos modelos, ao comparar resultados simulados e conhecidos, para visualizar PP e para gerar dados que não foram adquiridos [57], [58].

Para simular os modelos PP e MPP são utilizados métodos de Monte Carlo, tendo particular importância os métodos de *Markov Chain Monte Carlo* [57]. Desta forma pode-se prever a realização de pontos mais provável. Assim sendo, no próximo capítulo abordar-se-á a otimização de modelos com métodos de *Markov Chain Monte Carlo*.

Capítulo 5 Otimização através de métodos de Markov Chain Monte Carlo

Os métodos de simulação de Monte Carlo remontam a 1946, quando foram idealizados por Stan Ulam. Na sua essência trata-se de um método de simulação, através do qual se obtém um conjunto de amostras de uma função densidade de probabilidades conhecida, que são usadas para se obter uma aproximação da densidade pretendida, que pode ser complexa [61]. Recorrendo ao exemplo utilizado por *Metropolis* e *Ulam*, o cálculo da probabilidade de sucesso de um jogo de solitário envolve cálculos complexos, no entanto pode-se simular um número elevado de jogos e obter uma aproximação através da quantidade de sucessos [62].

A evolução dos métodos supramencionados são os *Markov Chain Monte Carlo* (MCMC). Estes tiram proveito das cadeias de Markov para que a simulação seja mais eficiente, através da extração de mais amostras de regiões mais importantes, isto é, de zonas de maiores probabilidades da distribuição alvo. Este tipo de métodos tem sido utilizado em inferência bayesiana, integração, mecânica estatística, simulação de sistemas físicos e otimização. De facto, foram tão importantes para as áreas da estatística, física, economia e engenharia, que uma das suas variantes, o algoritmo de Metropolis, foi considerada um dos dez algoritmos mais importantes de engenharia e ciência prática do século XX [61].

No âmbito da presente dissertação, assume especial importância a extensão aos métodos MCMC, elaborada por *Peter Green* em 1995, que permite a sua aplicação eficiente em casos cuja dimensão não é fixa, ao permitir fazer o “salto” entre subespaços, chamado *Reversible Jump Markov Chain Monte Carlo* (RJMCMC) [63]. Este tipo de método é importante na simulação de PPs, cujo número de pontos é desconhecido, como é o caso dos EVR.

Neste capítulo irá introduzir-se, brevemente, o conceito dos métodos de Monte Carlo, e o seu principal algoritmo. Em seguida, avançar-se-á para a abordagem MCMC, explicando-se os principais fundamentos de cadeias de Markov, e as vantagens da sua inclusão nos métodos anteriores. Explorar-se-á o algoritmo de Metropolis-Hastings, que serve de fundamento ao algoritmo RJMCMC, explicado a seguir. Por fim, irá apresentar-se o método de *simulated*

annealing, que em conjunto com os algoritmos anteriores, possibilita a otimização global de um modelo.

5.1. Monte Carlo

Os métodos de Monte Carlo surgem da necessidade de, por vezes, ser necessário calcular integrais de difícil resolução analítica, ou grandes somas, para se saber a probabilidade de um acontecimento ou esperanças. Assim, a ideia base consiste em extrair N amostras independentes e identicamente distribuídas, $\{x^i\}_{i=1}^N$, da distribuição alvo ($p(x)$), e calcular uma aproximação dessa distribuição, através de uma função discreta obtida por simulação (Equação (5.1)) [61].

$$p_N = \frac{1}{N} \sum_{i=1}^N \delta_{x^i}(x) \quad (5.1)$$

Quando o número de simulações é elevado, a lei forte dos grandes números permite afirmar que, com um grau de confiança elevado, se está perante uma boa aproximação da distribuição alvo. Além disso, se a variância da distribuição for finita, por exemplo σ^2 , então o teorema do limite central prova que existe convergência na distribuição do erro (Equação (5.2)), e que a variância do estimador é $\frac{\sigma^2}{N}$ [61], [64].

$$\sqrt{N}(\hat{p} - p) \xrightarrow[N \rightarrow \infty]{} N(0, \sigma^2) \quad (5.2)$$

Assim sendo, pode-se aproximar o integral $I(f)$ através de um conjunto de somas $I_N(f)$ (Equação (5.3)) [61], [64].

$$I_N(f) = \frac{1}{N} \sum_{i=1}^N f(x^i) \xrightarrow[N \rightarrow \infty]{} I(f) = \int_X f(x)p(x)dx \quad (5.3)$$

5.1.1. Amostragem de rejeição

A amostragem de rejeição é o algoritmo clássico de simulação de Monte Carlo. Consiste em simular uma distribuição $p(x)$, através de uma distribuição proposta, $q(x)$, multiplicada por uma constante M , de tal forma que $Mq(x)$ seja superior ou igual a $p(x)$ em todo o domínio, ou seja, $Mq(x) \geq p(x)$. Assim, extrai-se uma amostra de $q(x)$ que servirá para simular a distribuição alvo, mas esta não é aceite automaticamente. Para que isso aconteça, amostra-se, uniformemente, uma nova variável de decisão $u \sim U_{[0,1]}$, com a qual se compara a razão entre $p(x)$ e $Mq(x)$, o que faz com a quantidade de amostras aceites, seja proporcional à

probabilidade dessa amostra em $p(x)$ [61], [65]. Na Figura 5.1 pode-se consultar o algoritmo para o método Monte Carlo de amostragem de rejeição.

1. **Para** $i = 1$ até N
2. Amostrar $x^i \sim q(x)$
3. Amostrar $u \sim U_{[0,1]}$
4. **Se** $u < \frac{p(x^i)}{Mq(x^i)}$
5. Então aceita-se e incrementa-se o contador correspondente a essa variável
6. **Senão**, rejeita-se

Figura 5.1 - Algoritmo para o método Monte Carlo de amostragem de rejeição.

Esta é apenas uma forma de aplicação dos métodos de Monte Carlo. Apesar de existirem outros que fazem amostragens de formas mais eficientes, todos sofrem de limitações quando se utilizam modelo multivariados. Além disso, por vezes a escolha de M não é trivial [61].

5.2. Markov Chain Monte Carlo

A inclusão de cadeias de Markov nos métodos de Monte Carlo permite fazer amostragens mais eficientes, de tal forma que a corrente passe mais tempo nas regiões mais importantes da distribuição [61]. Para além disso, a principal limitação apontada aos métodos de Monte Carlo é a inadequabilidade para modelos multivariados, mas tal não se verifica nos métodos MCMC, devido à propriedade anterior [64].

5.2.1. Cadeias de Markov

Supondo o processo estocástico $x^{(i)}$, no espaço de estados finito, de dimensão s , e discreto $\chi = \{x_1, x_2, x_3, \dots, x_s\}$, considera-se uma cadeia de Markov se:

$$p(x^i | x^{i-1}, \dots, x^1) = K(x^i | x^{i-1}) \quad (5.4)$$

onde $K(\cdot | \cdot)$ é um *kernel* de transição, que no caso de espaços de estado com valores discretos e distribuição estacionária, é uma matriz de transição, independente do tempo. Assim se compreende que, nas cadeias de Markov, a única dependência existente é em relação ao estado temporal atual (x^{i-1}), ou por outras palavras, o futuro apenas depende do presente, sendo esta propriedade designada como propriedade de Markov [61], [65].

Após algum tempo, a cadeia converge para uma distribuição estacionária, ou invariante, (p), que é precisamente a distribuição que se pretende simular, independentemente do ponto inicial, e que é verdade apenas se obedecer à Equação (5.5). Assim, ao atingir este estado, a cadeia não abandona a distribuição invariante, ou seja, a distribuição é a mesma na iteração $i - 1$ e i , independentemente do novo estado [61], [64].

$$p_i = p_{i-1}K \quad (5.5)$$

Uma cadeia de Markov estacionária é aquela cuja distribuição condicional de x^i , dado x^{i-1} , é a mesma em qualquer instante [64]. No entanto, para que isso aconteça é necessário que o *kernel* obedeça à propriedade de irredutibilidade e à propriedade da aperiodicidade. Para que a propriedade da irredutibilidade se verifique, é necessário que estando num determinado estado, x , a probabilidade de visitar qualquer outro estado y , seja positiva ($p(y|x) > 0$). Para que uma cadeia de Markov seja aperiódica, não devem existir ciclos nos quais possa ficar presa [61], [65]. Estas propriedades são fundamentais no contexto de MCMC porque se uma cadeia de Markov for irredutível e aperiódica, então possui uma distribuição estacionária única [65].

Se a probabilidade de uma sequência de estados percorrida por uma cadeia de Markov for igual à probabilidade da mesma sequência percorrida em ordem inversa, então diz-se que a cadeia é reversível. Esta condição é suficiente para se poder afirmar que $p(x)$ é a distribuição invariante pretendida [61], [64]. Na Equação (5.6) encontra-se a condição de reversibilidade.

$$p(x^{(i)})K(x^{(i-1)}|x^{(i)}) = p(x^{(i-1)})K(x^{(i)}|x^{(i-1)}) \quad (5.6)$$

Partindo da relação anterior e somando ao longo de $x^{(i-1)}$ em ambos os membros fica [61]:

$$\begin{aligned} \sum_{x^{(i-1)}} p(x^{(i-1)})K(x^{(i)}|x^{(i-1)}) &= \sum_{x^{(i-1)}} p(x^{(i)})K(x^{(i-1)}|x^{(i)}) \\ &= \sum_{x^{(i-1)}} p(x^{(i-1)})K(x^{(i)}|x^{(i-1)}) = p(x^{(i)}), \end{aligned} \quad (5.7)$$

verificando-se, assim, a condição da Equação (5.5).

A condição de reversibilidade da Equação (5.7) está escrita relativamente ao caso discreto, porém, se o espaço de estados for contínuo, então deve-se verificar a Equação (5.8) [61].

$$p(x^{(i)}) = \int p(x^{(i-1)})K(x^{(i)}|x^{(i-1)})dx^{i-1} \quad (5.8)$$

Os modelos MCMC devem ser irredutíveis, aperiódicos e a distribuição alvo deve ser a distribuição invariante da cadeia de Markov, pelo que se deve ter o especial cuidado de obedecer à condição de reversibilidade, ao projetar os modelos [61].

5.2.2. Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings (MH) é o mais geral e conhecido dos algoritmos MCMC. A ideia base consiste na proposta de uma distribuição q , o denominado *kernel*, que permite amostrar os valores candidatos x^* , tendo em conta o estado atual x , com probabilidade $q(x^*|x)$. A distribuição gaussiana centrada no estado atual, com covariância Σ , definida como $q(x^*|x) = N(x^*|x, \Sigma)$, é uma distribuição comumente proposta, e nesse caso diz-se que se efetua um passeio aleatório gaussiano. No caso em que o *kernel* é $q(x^*|x) = q(x^*)$, então tem-se um amostrador independente [61], [65].

No entanto, não basta extrair um candidato para que a cadeia se mova para a sua localização. Assim, se $p(x)$ for a distribuição invariante alvo, é necessário calcular uma probabilidade de aceitação, A , com base numa razão de aceitação, α , através das expressões (5.9) e (5.10), respetivamente [65].

$$A = \min(1, \alpha) \tag{5.9}$$

$$\alpha = \frac{p(x^*)q(x^i|x^*)}{p(x^i)q(x^*|x^i)} \tag{5.10}$$

Se o α for maior que um, então a amostra é obrigatoriamente aceite, pois a sua probabilidade de aceitação é unitária. No entanto, se $0 \leq \alpha < 1$ ainda é possível mover a cadeia para locais menos prováveis, dependendo da comparação com uma variável amostrada de uma distribuição uniforme, U , o que faz com que o tempo despendido em cada estado seja proporcional a $p(x)$ [65]. Na Figura 5.2 pode-se consultar o algoritmo de MH.

Na Figura 5.3 pode-se verificar o efeito do número de iterações na simulação. Para tal, utilizou-se como distribuição alvo uma distribuição bimodal definida como $p(x) \propto 0,3e^{-0,2x^2} + 0,7e^{-0,2(x-10)^2}$, e como distribuição proposta uma distribuição gaussiana com desvio padrão 100. Executou-se o algoritmo de MH com 100, 1.000, 10.000 e 100.000 iterações, e tal como se pode verificar, quanto maior o número de iterações, melhor a simulação se assemelha à distribuição alvo.

1. Inicializar x^0
2. **Para** $i = 0$ até $N - 1$:
3. Amostrar $x^* \sim q(x^* | x^i)$
4. Amostrar $u \sim U_{[0,1]}$
5. Calcular $\alpha = \frac{p(x^*)q(x^i|x^*)}{p(x^i)q(x^*|x^i)}$
6. Calcular $A = \min(1, \alpha)$
7. Definir nova amostra $x^{i+1} = \begin{cases} x^* & \text{se } u < A \\ x^i & \text{se } u \geq A \end{cases}$

Figura 5.2 - Algoritmo de Metropolis-Hastings [61], [65].

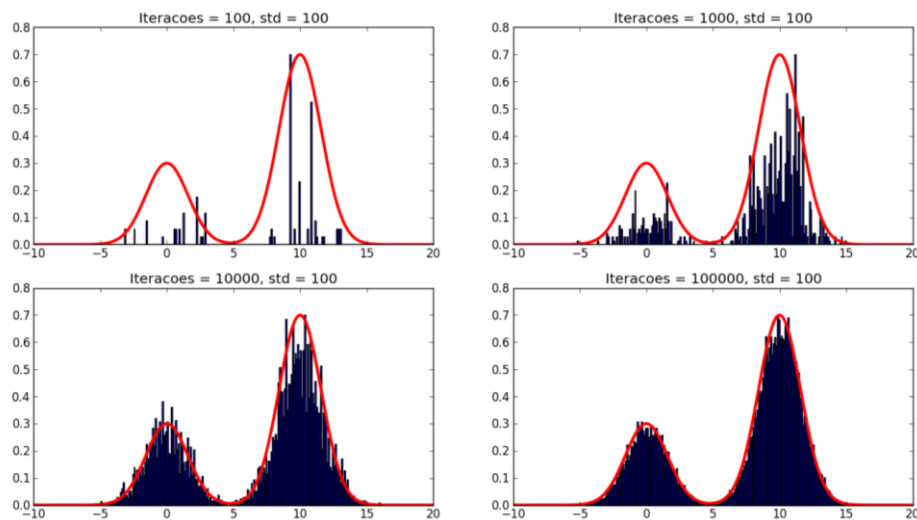


Figura 5.3 - Demonstração da influência do número de iterações na simulação. A linha vermelha representa a distribuição bimodal alvo, enquanto as barras azuis são o resultado da simulação.

Uma das vantagens do algoritmo de MH é a não obrigatoriedade de a distribuição alvo estar normalizada por uma constante Z , que muitas vezes é de difícil cálculo. Isto deve-se ao facto de a mesma se cancelar durante o cálculo de α , pois se se considerar $p'(x) = p(x)/Z$, então a equação (5.10) transforma-se na equação (5.11) [65].

$$\alpha = \frac{(p(x^*)/Z)q(x^i|x^*)}{(p(x^i)/Z)q(x^*|x^i)} \quad (5.11)$$

Uma outra propriedade do método verifica-se quando a distribuição proposta é simétrica ($q(x^*|x^i) = q(x^i|x^*)$), o que faz com que a probabilidade de aceitação fique reduzida à equação (5.12). Assim, o algoritmo passa a chamar-se apenas algoritmo de Metropolis, já que o

contributo de *Hastings* foi a inclusão da distribuição proposta no cálculo, como forma de correção no caso de a proposta ser assimétrica [61], [65].

$$A = \min\left(1, \frac{p(\mathbf{x}^*)}{p(\mathbf{x}^i)}\right) \quad (5.12)$$

5.2.2.1. Distribuições propostas

Para que uma distribuição proposta seja bem-sucedida é necessário que a sua probabilidade seja positiva em todo o espaço de estados da distribuição alvo. Este é outro motivo para a popularidade da distribuição gaussiana, já que é positiva em todo o espaço. Porém, ainda que esta propriedade se verifique, é necessário algum cuidado na proposta, uma vez que se não for adequada se pode concentrar numa moda e visitar as restantes poucas vezes, ou então rejeitar muitas propostas, ficando no mesmo estado muito tempo [65].

Caso exista dúvida quanto à proposta a utilizar, é possível fazer uma mistura de distribuições, desde que individualmente sigam a regra supramencionada. Assim, se se misturar K distribuições, com peso w_q na mistura, a distribuição proposta é [65]

$$q(\mathbf{x}^*|\mathbf{x}^i) = \sum_{k=1}^K w_k q_k(\mathbf{x}^*|\mathbf{x}^i) \quad (5.13)$$

Na Figura 5.4 pode-se observar o efeito que a distribuição proposta tem no resultado da simulação, particularmente quando se varia o desvio padrão da distribuição entre 0,1, 1, 10 e 100, para simular a distribuição bimodal definida como $p(x) \propto 0,3e^{-0,2x^2} + 0,7e^{-0,2(x-10)^2}$. Tal como se pode observar, se o desvio padrão for muito baixo (0,1) existe o risco de a cadeia se concentrar numa só moda, resultando numa simulação muito fraca, pois a probabilidade de sugerir uma amostra noutra região é muito baixa. Com desvio padrão igual a 1, verifica-se que a qualidade da simulação é superior, apesar de a cadeia ter estado demasiadas vezes na moda mais pequena. Com desvios padrões superiores, a qualidade da simulação melhora, o que demonstra que a distribuição proposta deve abranger a distribuição alvo, ou pelo menos ser “larga” o suficiente para permitir saltos para outras zonas.

5.2.2.2. Demonstração da condição de reversibilidade do algoritmo de Metropolis-Hastings

Como já foi referido, a condição de reversibilidade garante que se atingiu a distribuição estacionária, que é aquela que se pretende simular. Assim sendo, deve-se verificar a condição da Equação (5.6) para o algoritmo de MH.

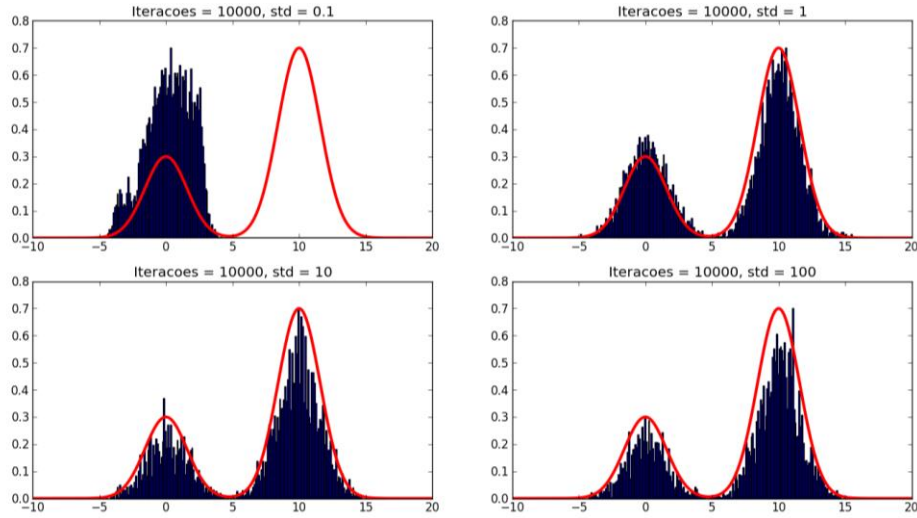


Figura 5.4 - Teste do efeito do desvio padrão da distribuição gaussiana proposta durante a simulação. A linha vermelha é a distribuição bimodal alvo, enquanto as barras azuis representam o resultado da simulação.

O *kernel*/de transição do algoritmo MH é

$$K(\mathbf{x}^*|\mathbf{x}^i) = \begin{cases} q(\mathbf{x}^*|\mathbf{x}^i)A(\mathbf{x}^*|\mathbf{x}^i), & \text{caso } \mathbf{x}^* \text{ seja aceite} \\ q(\mathbf{x}|\mathbf{x}) + \sum_{\mathbf{x}^* \neq \mathbf{x}} q(\mathbf{x}^*|\mathbf{x}^i) (1 - A(\mathbf{x}^*|\mathbf{x}^i)), & \text{caso contrário} \end{cases} \quad (5.14)$$

A probabilidade de uma proposta ser aceite é $A = \min(1, \alpha)$. Portanto se $\alpha < 1$ então $A(\mathbf{x}^*|\mathbf{x}^i) = \frac{p(\mathbf{x}^*)q(\mathbf{x}^i|\mathbf{x}^*)}{p(\mathbf{x}^i)q(\mathbf{x}^*|\mathbf{x}^i)}$ e $A(\mathbf{x}^i|\mathbf{x}^*) = 1$. Caso $\alpha > 1$, então $A(\mathbf{x}^*|\mathbf{x}^i) = 1$ e $A(\mathbf{x}^i|\mathbf{x}^*) = \frac{p(\mathbf{x}^i)q(\mathbf{x}^*|\mathbf{x}^i)}{p(\mathbf{x}^*)q(\mathbf{x}^i|\mathbf{x}^*)}$.

Assim sendo, caso se pretenda mudar de \mathbf{x}^i para \mathbf{x}^* , deve-se fazer a proposta e aceitá-la, assim

$$K(\mathbf{x}^*|\mathbf{x}^i) = q(\mathbf{x}^*|\mathbf{x}^i)A(\mathbf{x}^*|\mathbf{x}^i) = q(\mathbf{x}^*|\mathbf{x}^i) \frac{p(\mathbf{x}^*)q(\mathbf{x}^i|\mathbf{x}^*)}{p(\mathbf{x}^i)q(\mathbf{x}^*|\mathbf{x}^i)} = \frac{p(\mathbf{x}^*)}{p(\mathbf{x}^i)} q(\mathbf{x}^i|\mathbf{x}^*)$$

Continuando

$$p(\mathbf{x}^i)K(\mathbf{x}^*|\mathbf{x}^i) = p(\mathbf{x}^*)q(\mathbf{x}^i|\mathbf{x}^*)$$

Caso a mudança aconteça no sentido contrário, então $A = 1$, ficando

$$K(\mathbf{x}^i|\mathbf{x}^*) = q(\mathbf{x}^i|\mathbf{x}^*)A(\mathbf{x}^i|\mathbf{x}^*) = q(\mathbf{x}^i|\mathbf{x}^*)$$

Regressando à expressão anterior, e substituindo, fica-se com

$$p(\mathbf{x}^i)K(\mathbf{x}^*|\mathbf{x}^i) = p(\mathbf{x}^*)K(\mathbf{x}^i|\mathbf{x}^*)$$

Ficando, então, demonstrada a condição de reversibilidade do algoritmo de MH [65].

5.2.3. Reversible Jump Markov Chain Monte Carlo

O método RJMCMC [63], também conhecido como MCMC transdimensional, é uma evolução do algoritmo de MH que permite “saltar” entre subespaços de dimensões diferentes, sendo, portanto, utilizado para seleção de modelos, pois permite comparar modelos com diferente número de parâmetros. Assim, este método já foi utilizado para estimar o número de componentes de uma mistura de distribuições de número desconhecido, o número de neurónios de uma rede neuronal, o número de sinusóides num sinal com ruído, entre outros [61].

A dificuldade da aplicação do algoritmo MH, ou outros algoritmos MCMC, neste tipo de problemas é a impossibilidade de comparar densidades de objetos de diferentes dimensões na razão de aceitação. A solução passa por aumentar o modelo de menor dimensão através de novas variáveis aleatórias propostas com uma distribuição uniforme, de tal forma que ambos possam ser comparados sob uma determinada medida. Assim, a razão de aceitação passa a conter não só as densidades dos modelos, mas também a razão das medidas através da inclusão de um Jacobiano [61], [65]. Na Figura 5.5 pode-se analisar o princípio base do RJMCMC.



Figura 5.5 - Exemplificação do princípio do método *Reversible Jump Markov Chain Monte Carlo*. Duas funções densidade de probabilidade com dimensões diferentes não podem ser comparadas, mas adicionando uma nova variável proposta por uma distribuição uniforme à densidade de menor dimensão, e expandindo-a, já é possível fazer a comparação. Adaptado de [61].

Admitindo uma família finita de M modelos, $\{M_m: m = 1, \dots, N\}$, com propriedades desconhecidas $x_m \in \chi_m$, onde o espaço de estados é, por exemplo, $\chi_m = \mathbb{R}^{n_m}$, com n_m a dimensão do modelo m . Uma possibilidade para a escolha do modelo consiste em avaliar os modelos fazendo a amostragem dos parâmetros em todo o produto de espaços $\prod_{m=1}^M \chi_m$, porém esta estratégia é ineficiente, já que a maioria dos espaços não tem relevância e o tempo despendido não oferece vantagens. Assim, a abordagem do RJMCMC permite a diminuição do

espaço de procura, ao fazer a amostragem em $\chi \triangleq \bigcup_{m=1}^M \{m\} \times \chi_m$, tornando-se mais eficiente pois só se precisa de ter em conta os parâmetros do modelo atual [61], [65].

Para que se possa fazer o salto entre diferentes subespaços é necessário expandi-los para uma medida comum, para que sejam comparáveis. Assim sendo, se se pretender fazer o salto entre os subespaços χ_m e χ_n , é preciso expandir os espaços, com o recurso a variáveis auxiliares, para $\bar{\chi}_{m,n} \triangleq \chi_m \times U_{m,n}$ e $\bar{\chi}_{n,m} \triangleq \chi_n \times U_{n,m}$, respetivamente. Além disso, também deve existir uma função invertível (para assegurar a condição de reversibilidade) e diferenciável que faça a adaptação entre os espaços expandidos (equações (5.15) e (5.16)). É ainda necessário propor as distribuições $q_{n \rightarrow m}(\cdot | n, x_n)$ e $q_{m \rightarrow n}(\cdot | m, x_m)$ [61].

$$(x_m, u_{m,n}) = f_{n \rightarrow m}(x_n, u_{n,m}) = (f_{n \rightarrow m}^x(x_n, u_{n,m}), f_{n \rightarrow m}^u(x_n, u_{n,m})) \quad (5.15)$$

$$(x_n, u_{n,m}) = f_{m \rightarrow n}(f_{n \rightarrow m}(x_n, u_{n,m})) \quad (5.16)$$

Portanto, se o estado atual for (n, x_n) e se pretender mover para (m, x_m) deve-se gerar $u_{n,m} \sim q_{n \rightarrow m}(\cdot | n, x_n)$, ficando $x_m^* = f_{n \rightarrow m}^x(x_n, u_{n,m})$, e a razão de aceitação calcula-se através da Equação (5.17) [61].

$$A_{n \rightarrow m} = \min \left\{ 1, \frac{p(m, x_m^*)}{p(n, x_n)} \times \frac{q(n|m)}{q(m|n)} \times \frac{q_{m \rightarrow n}(u_{m,n} | m, x_m^*)}{q_{n \rightarrow m}(u_{n,m} | n, x_n)} \times J_{f_{n \rightarrow m}} \right\} \quad (5.17)$$

onde $J_{f_{n \rightarrow m}}$ é o jacobiano de $f_{n \rightarrow m}$ (equação (5.18)), quando estão em causa variáveis contínuas, e $q(\cdot | \cdot)$ representa a probabilidade de se escolher um *kernel*/de transição.

$$J_{f_{n \rightarrow m}} = \det \left| \frac{\partial f_{n \rightarrow m}(x_n, u_{n,m})}{\partial (x_n, u_{n,m})} \right| \quad (5.18)$$

As variáveis aleatórias que são geradas devem fazer com que exista uma bijeção entre $(x_m, u_{m,n})$ e $(x_n, u_{n,m})$, particularmente, o comprimento, l , dessas variáveis deve fazer com que a igualdade $l(x_m) + l(u_{m,n}) = l(x_n) + l(u_{n,m})$ seja verdadeira [63].

O método RJMCMC utiliza vários *kernels* de transição, também denominados por movimentos no contexto deste método, para percorrer os subespaços. Os movimentos devem ser reversíveis, e sempre que exista um movimento que permita mudar o modelo, deve existir outro que possibilite voltar atrás. Esta combinação de movimentos apresenta-se como a principal vantagem do método, uma vez que lhe confere a flexibilidade de utilizar múltiplos movimentos, tornando-o eficiente para fazer a seleção de modelos. Porém, também é o motivo para a sua difícil implementação, já que os movimentos têm que ser criados para cada aplicação, e a sua

construção também pode ser complexa [61][66]. Na Figura 5.6 pode-se consultar o algoritmo RJMCMC, onde t_i representa a probabilidade de se escolher o movimento i .

Entre os movimentos mais populares, referenciados em [24], [61], [63], encontram-se os movimentos *death*, *birth*, *split* e *merge*. Assim sendo, tanto o *birth* como o *split* estão associados ao aumento da dimensão do modelo, quer através da sugestão de uma nova variável, quer através da separação de uma variável em duas, respetivamente. Por outro lado, com o *death* escolhe-se uma variável que será retirada do modelo, enquanto o *merge* escolhe duas variáveis para serem fundidas. Geralmente, apenas com os movimentos *birth* e *death* é possível atingir-se a convergência, mas a adição de novos *kernels* de transição pode acelerar o processo [24]. Tomando como exemplo a seleção do número de componentes de um modelo de misturas, então o movimento *birth* corresponde à proposta de uma nova componente da mistura, enquanto o *death* removeria um componente uniformemente selecionado; por seu lado, o *split* separaria um componente em dois, enquanto o *merge* uniria dois componentes.

1. Inicializar (x_0, u_0)
2. **Para** $i = 0$ até $N - 1$:
3. Amostrar $u \sim U_{[0,1]}$
4. **Se** $(u \leq t_1)$
5. Então faz-se o movimento 1
6. **Senão**, se $(u \leq t_2)$, então faz-se o movimento 2
7. **Senão**, se $(u \leq t_3)$, então faz-se o movimento 3
8. ...
9. **Senão**, se $(u \leq t_n)$, então faz-se o movimento n
10. Amostragem de outros parâmetros.

Figura 5.6 - Algoritmo *Reversible Jump Markov Chain Monte Carlo*. u é a variável de decisão e t representa o movimento. Adaptado de [61].

5.3. *Simulated annealing* como forma de otimização

Uma forma possível de otimizar uma função através de métodos MCMC consiste em simular toda a função, restringindo o seu domínio se for infinito, e depois procurar o máximo. No entanto, esta abordagem é pouco eficiente porque muitas amostras não pertencem à moda da

função. Assim, surge o método de *simulated annealing*, no qual se procura estar mais tempo nos locais da vizinhança da moda [61].

Portanto, em 1983, *Kirkpatrick et al.* propuseram uma forma de otimização baseada no algoritmo de Metropolis, inspirada na mecânica estatística, cujo objetivo era calcular o máximo global de uma função [67]. No contexto do aquecimento e arrefecimento de materiais, a distribuição de Boltzmann (Equação (5.19)) traduz a probabilidade de um determinado estado, por exemplo o arranjo espacial dos átomos.

$$p(\mathbf{x}) \propto e^{\frac{-f(\mathbf{x})}{T}} \quad (5.19)$$

onde $f(\mathbf{x})$ é a energia do sistema, e T é a temperatura. À medida que a temperatura diminui, os átomos arranjam-se na forma de cristais, o que corresponde à energia mínima da distribuição de Boltzmann. Se o arrefecimento for lento, então a probabilidade de se formarem cristais é mais elevada, correspondendo a uma forma de energia mínima (máxima probabilidade), mas, caso o arrefecimento seja repentino, a organização dos átomos fica amorfa [67].

Na Equação (5.20) encontra-se a adaptação para a utilização da equação de Boltzmann para qualquer distribuição. Assim sendo, o método de arrefecimento simulado consiste na simulação de uma cadeia de Markov não homogénea, uma vez que em cada iteração a distribuição tem um expoente diferente [61].

$$p_i(\mathbf{x}) \propto e^{\frac{-f(\mathbf{x})}{T}} = (e^{-f(\mathbf{x})})^{\frac{1}{T}} = p(\mathbf{x})^{\frac{1}{T}} \quad (5.20)$$

A simulação é iniciada com temperatura elevada, que vai diminuindo ao logo do tempo, de acordo com um programa de arrefecimento estabelecido. Quando a temperatura é elevada, a cadeia pode percorrer qualquer região, à procura de modas, mas à medida que vai diminuindo, a probabilidade de aceitar as piores amostras também diminui, o que faz com que a cadeia se concentre numa moda, e evolua no sentido de maximizar a função [61], [65], [67].

Um dos problemas com o *simulated annealing* é a escolha do melhor programa de arrefecimento, já que se este for muito rápido pode resultar numa solução local, enquanto que se for demasiado lento torna o processo demorado. Na prática, é utilizado com frequência o arrefecimento exponencial (Equação (5.21)), mas a velocidade de arrefecimento é dependente da aplicação, pelo que deve ser determinada após algumas experiências [61], [65].

$$T_k = T_0 C^k \quad (5.21)$$

onde T_k é a temperatura na iteração k , T_0 é a temperatura inicial e C é a taxa de arrefecimento.

Apesar de se ter falado sempre do *simulated annealing* com o algoritmo de MH, este também pode ser aplicado com o RJMCMC, de uma forma semelhante [24]. Na Figura 5.7 encontra-se o algoritmo de MH, alterado para executar a otimização global.

1. Inicializar x^0
2. **Para** $i = 0$ até $N - 1$:
3. Amostrar $x^* \sim q(x^* | x^i)$
4. Amostrar $u \sim U_{[0,1]}$
5. Calcular $\alpha = \frac{p(x^*)^{\frac{1}{T_i}} q(x | x^*)}{p(x)^{\frac{1}{T_i}} q(x^* | x)}$
6. Calcular $A = \min(1, \alpha)$
7. Definir nova amostra $x^{i+1} = \begin{cases} x^* & \text{se } u < A \\ x^i & \text{se } u \geq A \end{cases}$
8. Atualizar T_i de acordo com o programa de arrefecimento.

Figura 5.7 - Algoritmo de Metropolis-Hastings, alterado para fazer otimização global, através de *simulated annealing*. Adaptado de [61].

Na Figura 5.8 pode-se observar o efeito da taxa de arrefecimento para descobrir o máximo global de uma distribuição bimodal definida como $p(x) \propto 0,3e^{-0,2x^2} + 0,7e^{-0,2(x-10)^2}$, utilizando como distribuição proposta uma gaussiana com desvio padrão igual a 100. Quanto mais elevado for C , mais lento é o arrefecimento. Assim, quando o arrefecimento é o mais lento, o método consegue descobrir o máximo global, mas também se verifica que despendeu algum tempo noutros pontos, inclusive na moda mais pequena. Para $C = 0,8$, o algoritmo foi capaz de descobrir um resultado próximo do máximo global, gastando a maioria do seu tempo aí. No que se refere a arrefecimentos mais rápidos, não só não foi possível fazer a otimização global da distribuição, como ainda ficou retido muito tempo noutros locais, inclusive em regiões próximas da moda mais baixa, comprovando aquilo que já foi referido acerca da importância da escolha do programa e taxa de arrefecimento.

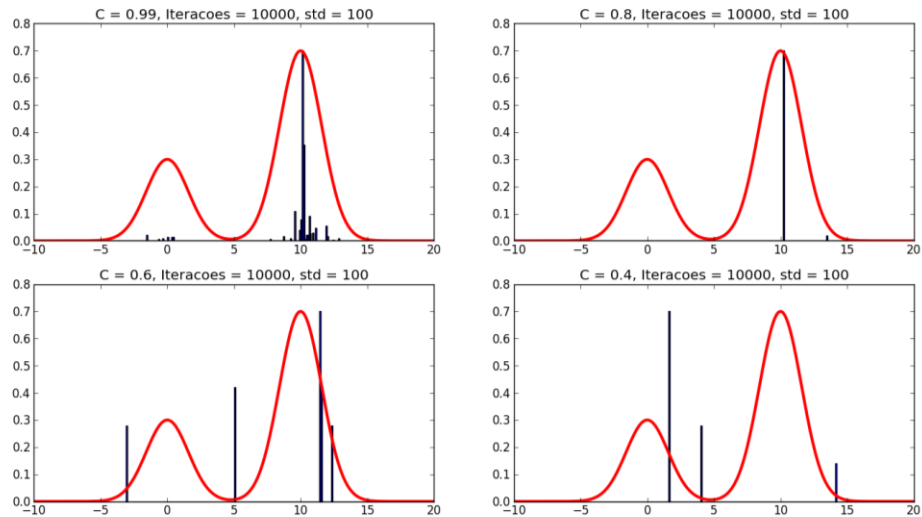


Figura 5.8 - Efeito da taxa de arrefecimento no método de *simulated annealing*. A vermelho encontra-se a distribuição bimodal alvo, enquanto as barras azuis são o resultado da simulação.

5.4. Sumário

Alguns problemas são tão complexos que são analiticamente intratáveis, portanto uma das formas de os abordar é através de métodos de simulação. Entre esses, encontram-se os métodos de Monte Carlo. Nestes, propõe-se uma distribuição, da qual se vai extraindo amostras, que são aceites segundo algum critério, e no final obtém-se a distribuição subjacente.

Se se incorporar cadeias de Markov, então está-se perante um método MCMC, tendo-se especificado o algoritmo de MH. Nas cadeias de Markov, existe um *kernel* de transição de estados, de tal forma que para se calcular a probabilidade de se deslocar para um novo estado, basta saber o estado anterior. Após algumas transições, a cadeia tende para a sua distribuição invariante, que é aquela que realmente governa as suas transições. Assim, ao incorporar-se estas propriedades, os métodos MCMC sugerem mais amostras de regiões de maior probabilidade, tornando-os mais precisos e eficientes.

Os algoritmos anteriores são aplicados em problemas de dimensão constante. No caso dos EVR, pretende-se descobrir qual é o conjunto de candidatos mais provável, o que envolve ir mudando a dimensão do problema, à medida que se removem ou adicionam novos candidatos. Assim, surge o método RJMCMC, como extensão ao MCMC, que permite abordar esse tipo de problema, ao utilizar vários *kernels*, fazendo a adaptação entre os espaços.

Descobrir a configuração mais provável de EVR é um problema de otimização global. Para tal, introduziu-se o conceito de *simulated annealing*. Este consiste em amostrar novas

propostas, a partir da distribuição proposta elevada a um parâmetro de temperatura, o qual se vai reduzindo, com o tempo. No final, o amostrador fica “preso” na região mais provável do modelo, correspondendo à configuração que o maximiza, globalmente.

Capítulo 6 Reconhecimento de padrões

O reconhecimento de padrões é um ramo do campo de aprendizagem máquina, cujo objetivo é a classificação automática de um objeto em categorias, ou classes. O objeto pode ser um sinal, uma imagem, caracteres ou outro tipo de dados que necessite de categorização. Assim, para que a classificação seja bem-sucedida é necessário identificar padrões nos dados, que sejam, tanto quanto possível, bons indicadores de cada classe [68].

Quando se pretende identificar os padrões é necessário que os algoritmos sejam capazes de os aprender ou perceber, assim existem algoritmos de aprendizagem supervisionada, não-supervisionada e semi-supervisionada. Na aprendizagem supervisionada, é imperativo que exista, *a priori*, um conjunto de dados de treino cuja classe é conhecida, de tal forma que o algoritmo seja capaz de ser treinado e aprender padrões. Por outro lado, na aprendizagem não-supervisionada não existem dados de treino, por isso os algoritmos recebem o objeto de análise e devem ser capazes de perceber similaridades entre as características e agrupá-las. Já a aprendizagem semi-supervisionada fica num ponto intermédio entre as duas anteriores, já que tem um conjunto de dados de treino, em que só uma parte (geralmente pequena) tem a respetiva classe associada. Os algoritmos de reconhecimento de padrões que utilizam aprendizagem supervisionada são chamados classificadores, enquanto os que seguem abordagens não-supervisionadas são denominados por algoritmos de *clustering* [68].

Os conceitos supramencionados podem ser utilizados para segmentação de imagens, ao atribuir-se uma classe a cada *voxel*. Assim sendo, pode-se segmentar uma região de interesse no cérebro, onde a deteção de EVR seja mais relevante, aplicando-se o algoritmo de deteção automática dessas estruturas somente nessa região.

Neste capítulo começar-se-á por introduzir os conceitos globais de reconhecimento de padrões para abordagens supervisionadas. Posteriormente, irá expor-se os principais conceitos de *Random Decision Trees*, ou *Decision Trees* (DTs). Estes irão fundamentar o último tópico, que aborda as *Random Decision Forests*, ou *Random Forests* (RFs).

6.1. Conceitos globais de reconhecimento de padrões

Nesta secção irá apresentar-se alguns conceitos de reconhecimento de padrões necessários para a compreensão das RFs, que tem o maior foco na presente dissertação. Assim, expõe-se o conceito de classificador, características e as várias estratégias de avaliação e validação dos métodos de reconhecimentos de padrões e aprendizagem máquina.

6.1.1. Classificador

Um classificador é um método que categoriza/classifica um objeto com base em padrões encontrados neste. Para isso, deve ser treinado tendo como entrada um conjunto de dados cuja classe é conhecida, de tal forma que consiga fazer o mapeamento das características de entrada, \mathbf{x} , para o conjunto de saída $c \in \{1, \dots, C\}$, sendo c a classe. Assim sendo, um classificador pode ser representado por uma função $c = f(\mathbf{x})$, cuja fase de treino implica aprender e otimizar f , de tal forma que possa ser aplicável a novos dados, desconhecidos. A otimização do classificador deve procurar uma forma de separar as classes [65].

Se o processo de classificação apenas envolver a distinção entre duas classes, então trata-se de um problema de classificação binária. Se, por outro lado, se pretende que o resultado compreenda várias classes, então trata-se de classificação multi-classe [65].

É desejável que um classificador tenha boa capacidade de generalização, isto é, deve ser capaz de classificar corretamente os dados de teste, que são diferentes dos que foram utilizados para o treino. Portanto, na fase de treino a escolha de parâmetros deve ser cuidada, para evitar o problema de *overfitting*, ou seja, deve-se evitar que o método aprenda tão bem os dados de treino que apenas consiga distinguir os novos casos que sejam muito semelhantes àqueles que foram utilizados na otimização, o que representa um problema se existirem *outliers* nos dados de treino [65], [68].

Entre os exemplos de classificadores pode-se apontar as *Support Vector Machines*, as Redes Neurais ou as RFs.

6.1.2. Características

As características são quantidades mensuráveis que capturam propriedades dos objetos, de tal forma que sejam distintas entre as classes. Num processo de classificação podem ser utilizadas várias (n) características, que formam um vetor $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ [68].

Muitas vezes, não se pode utilizar somente o objeto adquirido para a classificação, pois este pode ter grande correlação entre os seus constituintes, por exemplo nos sulcos mais

estreitos do córtex cerebral existe o efeito de volume parcial nas imagens de RM, o que pode fazer com que o LCR e a SC tenham intensidades próximas, nesses *voxels*. Nesse sentido, deve-se procurar a melhor forma de captar informação distinta entre classes, sendo que, geralmente, a utilização de características dependentes da utilização pretendida costuma obter melhores resultados, por exemplo, se se souber que uma classe de um objeto de uma imagem tem um determinado tipo de textura, então deve-se incluir essa informação no classificador, através de características de textura. Assim, em análise de imagem pode-se calcular características de textura (média, detecção de bordas, matriz de co-ocorrência, entre outras), de vizinhança do *pixel/voxel*, a própria intensidade das imagens, resultados de filtros, análise de formas, entre outras [68].

6.1.3. Avaliação do modelo de classificação

O classificador é treinado com dados cujas classes são conhecidas, mas para avaliar o sucesso da classificação também é necessário saber a classe de cada ponto de teste. Assim, como esse tipo de dados é limitado, deve-se escolher uma estratégia de treino e teste, utilizando os pontos disponíveis, que permita avaliar a classificação e o impacto da variação dos parâmetros do classificador e das características [68], [69].

O método de *Resubstitution* utiliza todos os dados disponíveis para treino e para teste. Porém, a reutilização dos dados resulta numa avaliação otimista e pouco realista [68].

Uma outra forma de avaliação dos modelos mais credível e utilizada, é a validação cruzada, que consiste na divisão dos dados disponíveis num conjunto de treino e de teste. Dentro deste tipo existem as abordagens *Holdout*, *K-fold* e *Leave-one-out* [68], [69].

Assim sendo, na estratégia *Holdout* o conjunto de dados disponíveis é dividido nos subconjuntos mutuamente exclusivos de treino e teste, sendo, geralmente, o número de elementos do conjunto de treino superior ao de teste, em prol da capacidade de generalização. Esta estratégia sofre de vários problemas, pois o classificador não tem acesso a todos os dados disponíveis, o conjunto de treino e de teste ficam, ambos, reduzidos e a escolha da proporção da divisão também é um problema de difícil resolução. Isto é particularmente grave se, após a divisão aleatória, uma classe ficar com poucos pontos de treino, prejudicando a generalização, mas se uma classe ficar com poucos pontos de teste, a avaliação também perde confiança. Em resumo, a abordagem *Holdout* faz uma utilização ineficiente dos dados disponíveis [68], [69].

Já a abordagem *K-fold* divide os dados em K conjuntos, mutuamente exclusivos, com igual número de elementos. Assim, o classificador é treinado e testado K vezes, ou seja, cada

treino utiliza $K - 1$ conjuntos, e o teste é realizado no que ficou de fora. Apesar de se atenuar alguns problemas da abordagem *Holdout*, o classificador pode, ainda assim, não ter acesso aos exemplos de treino necessários [68], [69].

Por seu lado, a abordagem *leave-one-out* divide os dados em tantos subconjuntos quantas as amostras disponíveis (N), por exemplo, se existirem 20 imagens, cada imagem é uma amostra, ou seja, um subconjunto. Assim, o treino utiliza $N - 1$ amostras, sendo testado na que não foi utilizada, e no final calcula-se a média dos N testes. Esta abordagem é aquela que disponibiliza o maior conjunto de dados de treino e teste ao classificador, mas a sua exigência computacional é mais elevada [68], [69].

Uma forma de tentar equilibrar os exemplos dos conjuntos de dados de treino e de teste da avaliação cruzada é a utilização de estratificação. Assim sendo, a divisão dos subconjuntos é feita de tal forma que a proporção de cada classe em cada um deles seja a mesma do conjunto total, evitando que uma das classes possa não estar presente [68], [69]. De facto, *Kohavi*, em [69], recomenda utilizar validação cruzada *K-fold*, com $K = 10$, e estratificação.

As abordagens supramencionadas referem-se à avaliação do grau de sucesso da classificação, mas a avaliação do melhor modelo para o classificador implica avaliar quais os hiperparâmetros que promovem uma melhor capacidade de generalização do método. Os hiperparâmetros são os parâmetros que devem ser especificados para cada método, e que não são otimizados durante o processo de treino. Uma das formas de se avaliar os hiperparâmetros denomina-se *grid search*, e consiste na avaliação do classificador para diferentes combinações desses parâmetros [70].

6.2. Fundamentos de *Random Forests* enquanto classificador

As RFs resultam da conjugação de várias DTs, e podem ser utilizadas em problemas de regressão, aprendizagem semi-supervisionada e classificação, sendo este último o foco do presente capítulo. As DTs são consideradas classificadores fracos, isto é, a sua capacidade de classificação é apenas um pouco superior à da classificação aleatória. Porém, pode-se treinar vários destes classificadores, atribuindo diferentes pesos aos dados, ou seja, para treinar o classificador fraco seguinte deve-se dar mais peso aos dados que foram mal “aprendidos” pelos anteriores, numa abordagem denominada por *boosting*. Desta forma, a combinação linear de vários classificadores fracos, treinados com diferentes dados, resulta num classificador forte, que

se correlaciona fortemente com a verdadeira classificação, uma vez que demonstram grande precisão e capacidade de generalização. [71]

As principais vantagens das RFs, enquanto classificador, são a sua capacidade de serem aplicadas em problemas multi-classe, o seu resultado ser probabilístico, a sua generalização para novos dados, a sua eficiência (por ser paralelizável) e a facilidade de interpretação [65], [71].

6.2.1. *Decision Trees*

Uma árvore é um grafo com estrutura hierárquica, o que significa que cada nó, excetuando o primeiro, é precedido por um, e só um, vértice. Para além disso, não podem existir ciclos entre os nós. Fazendo a analogia com uma árvore, este tipo de grafos começa por um só nó que representa a raiz, e que se divide em dois ramos que terminam em novos nós. Estes nós, os nós internos, dividem-se novamente em dois ramos, e assim sucessivamente até se atingir um ponto em que já não se dividem mais. Os nós que já não dão origem a mais vértices são os terminais, designados por folhas. Este tipo de árvore, em que cada nó interno tem dois vértices de saída, é denominada árvore binária (Figura 6.1(a)) [71].

Os nós correspondem a avaliações relativas a determinados *thresholds*, com os quais se compara os atributos do objeto sob análise, e que resulta na escolha de um vértice que deve ser seguido, até se atingir uma folha, onde se faz a classificação. Portanto, pode-se imaginar uma DT como uma sequência de perguntas que resultam numa conclusão acerca de um dado assunto [65], [71]. Na Figura 6.1(b) está um exemplo de utilização para classificar animais.

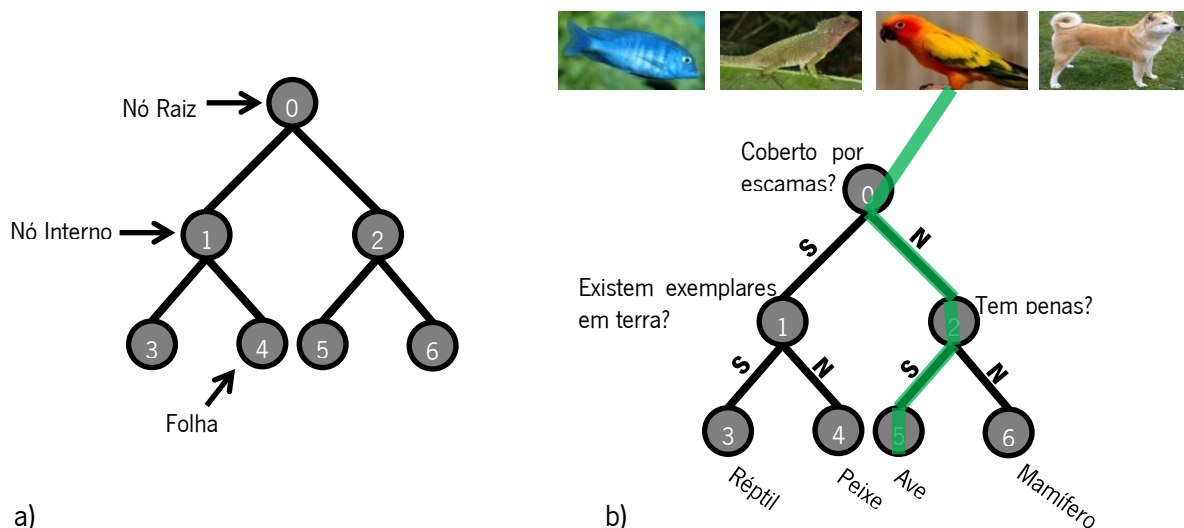


Figura 6.1 – *Decision Tree*. Em a) pode-se observar a estrutura hierárquica de uma DT, enquanto em b) se demonstra o princípio de classificação numa DT, utilizando como exemplo a classificação de uma ave.

Os parâmetros das árvores, como os *thresholds* a serem aplicados e as características a serem avaliadas em cada nó, devem ser aprendidos de forma supervisionada, recorrendo a um conjunto de dados de treino [71].

6.2.1.1. Modelo matemático das *Decision Trees*

As DTs avaliam, em cada nó, um conjunto de características do objeto sob análise. Considere-se que o objeto de análise é uma imagem de RM, então cada *voxel*, ou ponto, pode ser representado por um vetor de características de dimensão d , escrito sob a forma de $\mathbf{v} = (x_1, x_2, \dots, x_d) \in F$. Uma vez que em cada nó não são testadas todas as características, estas não tem que ser todas extraídas *a priori*, ou seja, podem ser extraídas à medida que o ponto percorre a árvore, o que torna as DTs mais eficientes em relação a outros métodos [71].

As características são sujeitas a um teste em cada nó, isto é, são sujeitas à decisão de uma função de teste (Equação (6.1)). O resultado é 0 ou 1, dependendo da condição aplicada, e condiciona o vértice que \mathbf{v} vai seguir.

$$h(\mathbf{v}, \theta_j): F \times \mathcal{T} \rightarrow \{0,1\} \quad (6.1)$$

onde $\theta_j \in \mathcal{T}$ representa os parâmetros da função de teste, no nó j [71].

Para otimizar os parâmetros das referidas funções é necessário uma fase em que se avalia o resultado da divisão dos pontos, durante o treino. Assim, cada ponto de treino está associado à sua classe y , que é conhecida, escrevendo-se (\mathbf{v}, y) . A raiz, ou nó 0, tem acesso a todos os pontos do conjunto de treino, S_0 , que se vai dividindo em subconjuntos à medida que vão percorrendo os nós. Desta forma, o subconjunto de treino do nó 1 é S_1 , e os subconjuntos que resultam da otimização e aplicação da função de teste são S_1^E , se for para o nó da esquerda, e S_1^D , se for para o nó da direita, resultando nos subconjuntos S_3 e S_4 do exemplo da Figura 6.1(a), respetivamente. As DTs binárias obedecem às seguintes propriedades [71]:

$$S_j = S_j^E \cup S_j^D$$

$$S_j^E \cap S_j^D = \emptyset$$

$$S_j^E = S_{2j+1}$$

$$S_j^D = S_{2j+2}$$

6.2.1.2. Treino e teste das *Decision Trees*

Durante o treino, a divisão em cada nó é otimizada através de função de separação dos pontos do subconjunto S_j em S_j^E e S_j^D , tendo em conta as características. Para isso, deve-se fazer a maximização de uma função objetivo (Equação (6.2)) [71].

$$\boldsymbol{\theta}_j^* = \arg \max_{\boldsymbol{\theta}_j \in \mathcal{T}} I_j \quad (6.2)$$

com

$$I_j = I(S_j, S_j^E, S_j^D, \boldsymbol{\theta}_j)$$

$$S_j^E = \{(\mathbf{v}, y) \in S_j | h(\mathbf{v}, \boldsymbol{\theta}_j) = 0\}$$

$$S_j^D = \{(\mathbf{v}, y) \in S_j | h(\mathbf{v}, \boldsymbol{\theta}_j) = 1\}$$

A escolha do tipo de função de separação pode ter impacto na capacidade de generalização do classificador. Podem-se escolher funções lineares, que tomam a forma de hiperplanos em dimensões elevadas, ou funções não lineares, que permitem mais graus de liberdade. Porém, é possível incluir formas mais complexas de separação, como a utilização de *Support Vector Machines* nos nós [71].

Durante o treino também se otimiza a estrutura da árvore. Assim sendo, pode-se deixar a árvore crescer até ao máximo, ou impor um valor à fase de otimização que impeça que esta fique muito profunda. Intuitivamente, poder-se-ia pensar que uma árvore totalmente crescida teria maior poder de discriminação, mas também se pode dar o caso do número de pontos que chegam às folhas ser tão baixo que ocorra um problema de *overfitting*, prejudicando a capacidade de generalização [71]. Entre os critérios de paragem de crescimento, ou seja, de divisão dos nós, pode-se incluir a avaliação da redução do custo de separação, um parâmetro que estipule a profundidade máxima permitida às DTs, a avaliação da homogeneidade das distribuições dos nós filhos (se os pontos forem todos da mesma classe, então a distribuição é pura) e o número de pontos em cada um dos nós filhos, que não deve ser demasiado reduzido [65].

A função objetivo utilizada durante o treino é importante, na medida em que influencia o caminho que os pontos de treino tomam e, conseqüentemente, o caminho que os pontos de teste irão seguir durante o teste, refletindo-se na precisão da classificação. Assim, tomam especial importância na construção das funções os conceitos de entropia e ganho de informação. Imagine-se que num nó existem pontos de três classes, mas, após a utilização de uma função de separação linear, em cada nó filho só existem pontos de duas classes, então neste caso a classificação fica mais fácil se forem utilizados os nós filhos. Portanto, a redução da incerteza é uma abordagem positiva para melhorar a classificação, e a avaliação do grau de incerteza pode ser feita com a utilização da entropia e do ganho de informação [71].

Ao fazer a divisão supramencionada reduz-se o número de classes nos nós filhos, ficando mais puros, o que diminui a entropia (Equação (6.3)). Por outro lado, a informação

contida em cada nó aumenta o poder discriminativo da DT, isto é, aumenta o ganho de informação (Equação (6.4)). Se, por outro lado, se dividir os pontos pelos nós filhos de tal forma que num existam duas classes e no outro exista apenas uma, então o ganho de informação aumenta, o que demonstra o potencial de se utilizar esta medida como função objetivo a maximizar [71]. Também o coeficiente de Gini pode ser utilizado como critério de avaliação da divisão de um nó, pois avalia o erro futuro da classificação (Equação (6.5)) [65]. Na Figura 6.2 encontra-se o exemplo utilizado, com as três classes.

$$H(S) = - \sum_{c \in C} p(c) \log(p(c)) \quad (6.3)$$

$$I = H(S) - \sum_{i \in \{E,D\}} \frac{|S^i|}{|S|} H(S^i) \quad (6.4)$$

$$G = 1 - \sum_{c=1}^C p(c)^2 \quad (6.5)$$

onde $p(c)$ representa a distribuição da classe c , construída com base nos pontos do conjunto S .

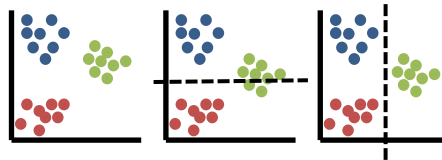


Figura 6.2 - Demonstração da divisão dos pontos de duas maneiras diferentes num nó de um *Decision Tree*. Mais à esquerda observa-se todos os pontos do nó pai; no centro encontra-se uma possível divisão dos pontos para os dois nós filhos, e conclui-se que a classificação vai ser mais precisa, particularmente em relação às classes azul e vermelha; mais à direita encontra-se uma divisão que favorece a classificação da classe verde. Da esquerda para a direita há diminuição da entropia, e aumento do ganho de informação.

Durante o treino, os pontos são divididos pelos nós até que nas folhas apenas restem subconjuntos dos pontos iniciais. Durante o teste, cada ponto \mathbf{v} percorre os nós da DT até atingir uma folha, que, à partida, contém pontos de treino semelhantes. Caso se utilize a árvore como classificador, então na folha o ponto é sujeito a um estimador que atua sobre a probabilidade condicionada de o mesmo pertencer a cada uma das classes ($p(c|\mathbf{v})$ com $c \in c_k$), calculada com base nas estatísticas dos pontos de treino aí presentes. Um estimador possível é *Maximum a-Posteriori*, que seleciona a classe com maior probabilidade condicionada, estimada através de $c^* = \arg \max_c p(c|\mathbf{v})$ [71].

No treino é adicionada alguma aleatoriedade às árvores, quer seja através da utilização de amostras de treino aleatórias extraídas do conjunto de treino, que depois são repostas, num

processo denominado por *bagging*, quer seja através da otimização dos nós de forma aleatória [71].

A otimização dos nós de forma aleatória torna o treino mais eficiente, o que permite que se utilize todos os pontos exemplo e que exista a propriedade de maximização da margem, quando se conjugam várias árvores para criar a RF. Assim, em vez de se otimizar todo o conjunto de parâmetros, \mathcal{T} , apenas se otimiza um subconjunto, $\mathcal{T}_j \in \mathcal{T}$, escolhido aleatoriamente. A quantidade de aleatoriedade é controlada pela razão $|\mathcal{T}_j|/|\mathcal{T}|$; surge, então, a necessidade de se estabelecer um novo parâmetro que controle o grau de aleatoriedade, $\rho = |\mathcal{T}_j|/|\mathcal{T}|$, que é, geralmente, igual para todos os nós. Portanto, se $\rho = 1$, então não existe aleatoriedade, e se $\rho = 0$ está-se perante aleatoriedade total, ou seja, não existe otimização [71].

6.2.2. Random Forests

As RFs são constituídas por T DTs, com $t \in \{1, 2, \dots, T\}$, que obedecem às características discutidas anteriormente. Assim sendo, o parâmetro ρ controla, não só a aleatoriedade de cada DT individual, mas também da RF, o que faz com que todas as árvores sejam diferentes, diminuindo a correlação entre elas e aumentando a capacidade de generalização da floresta [71].

Cada árvore é treinada e testada independentemente, o que faz com que o processo seja paralelizável e eficiente. Portanto, durante o teste, cada \mathbf{v} é colocado na raiz de todas as DTs, percorrendo-as até atingir as folhas. No final, o resultado das DTs deve ser combinado, de maneira a que se obtenha uma classificação global para a RF; para isso, pode-se calcular a média dos resultados das árvores (Equação (6.6)), ou a sua multiplicação (Equação (6.7)) [71].

$$p(c|\mathbf{v}) = \frac{1}{T} \sum_{t=1}^T p_t(c|\mathbf{v}) \quad (6.6)$$

$$p(c|\mathbf{v}) = \frac{1}{Z} \prod_{t=1}^T p_t(c|\mathbf{v}) \quad (6.7)$$

onde Z é uma função de normalização das probabilidades.

O problema da classificação consiste em, dado um ponto \mathbf{v} , descobrir a classe $c \in \mathcal{C} = \{c_k\}$ cuja $p(c|\mathbf{v})$ é máxima, utilizando uma RF treinada com pontos cuja classe era conhecida, ou seja, (\mathbf{v}, c) . Também com as florestas, é necessário fazer a otimização dos nós de forma semelhante às DTs, isto, utilizando a entropia e o ganho de informação, ou o coeficiente de Gini,

com as Equações (6.2), (6.3), (6.4) e (6.5). Um problema comum em classificação de imagens é o problema das classes não balanceadas, quando existem classes com presença muito superior, relativamente a outras, no conjunto de treino, o que pode afetar a estrutura e precisão das RFs. De forma a mitigar o problema, pode-se atribuir pesos diferentes às classes ou fazer reamostragem das que se encontrem em desvantagem [71].

6.2.2.1. Efeito dos parâmetros das *Random Forests*

Os parâmetros que mais afetam as RFs são: a profundidade máxima das DTs (D), a quantidade de aleatoriedade (ρ), o número de árvores (T), a função de separação, a função objetivo a ser otimizada e as características escolhidas para classificar o objeto sob estudo [71].

O número de árvores, T , ou dimensão da floresta afeta a capacidade de generalização do classificador. Quando se utiliza apenas uma DT a função de separação estabelece um limite rígido de separação entre os pontos das classes, o que não é desejável, pois os pontos de teste são diferentes do treino, e aqueles que ficarem muito próximo da separação das classes podem ser erroneamente classificados. Ao se utilizar uma RF, cada árvore nela contida é única, o que faz com que deixe de existir um limite de separação, passando a existir uma região, ou seja, a confiança junto das regiões dos pontos de treino é elevada, mas vai diminuindo à medida que se afasta deles. Desta forma, vai suavizando as probabilidades *a posteriori* das classes, o que é positivo para a capacidade de generalização do classificador. A precisão do teste aumenta à medida que se amplia T , mas a dado momento a melhoria é tão reduzida que não compensa a perda de eficiência [71].

A profundidade máxima das árvores, D , está diretamente relacionada com a confiança da classificação, numa relação de proporcionalidade direta. Assim, se as DTs forem muito pequenas a confiança é baixa, mas aumenta à medida que se aumenta a profundidade de crescimento. Porém, florestas com D elevado correm o risco de *overfitting*, isto é, podem-se adaptar tanto aos dados de treino que até os pontos provenientes de ruído são incluídos fortemente no modelo, o que prejudica a generalização e a precisão, quando são apresentados pontos diferentes, durante o teste. Aumentar o número de árvores da RF diminui este efeito, mas não o remove totalmente. Para além desses problemas, ao se permitir elevadas profundidades pode acontecer que as árvores não fiquem balanceadas [71]. Também é possível, num passo posterior, podar as árvores que cresceram em demasia, como forma de reduzir o *overfitting* [65]. Na Figura 6.3(a) demonstra-se o efeito do *overfitting* com o treino a adaptar o modelo para pontos que são *outliers* das classes, uma vez que já se encontram nas regiões de outras classes

(círculos brancos). Os pontos de teste que forem semelhantes a esses pontos podem ser mal classificados.

Relativamente às funções de separação, utilizar separação linear de tal forma que as retas sejam paralelas aos eixos cartesianos torna a classificação muito eficiente, mas pode criar artefactos na forma de blocos de elevada confiança, o que não é positivo para a capacidade de generalização e precisão (Figura 6.3(b)). Já a utilização de funções não lineares, por exemplo em estrutura cônica, corrige esse problema ao ser capaz de se adaptar melhor à distribuição dos dados de treino, mas exige maior carga computacional (Figura 6.3(c)). Uma alternativa intermédia seria a utilização de funções lineares orientadas, ou seja, que não seriam paralelas aos eixos [71].

A aleatoriedade durante a otimização, ρ , ajuda a diminuir os artefacto de blocos das funções lineares paralelas aos eixos, uma vez que diminui a confiança de uma forma geral em todo o espaço (Figura 6.3(d)). Já as funções não lineares têm maior número de parâmetros, e a adição de aleatoriedade faz com que só alguns sejam otimizados, tornando a utilização destas funções mais eficiente. A utilização de um elevado grau de aleatoriedade para reduzir o número de parâmetros a otimizar das funções não lineares pode fazer com que a confiança geral fique demasiado baixa, mas pode-se compensar ao aumentar a profundidade das árvores. Uma outra vantagem da aleatoriedade é o facto de tornar as árvores de uma RF únicas e diferentes entre si, o que aumenta a capacidade de generalização [71].

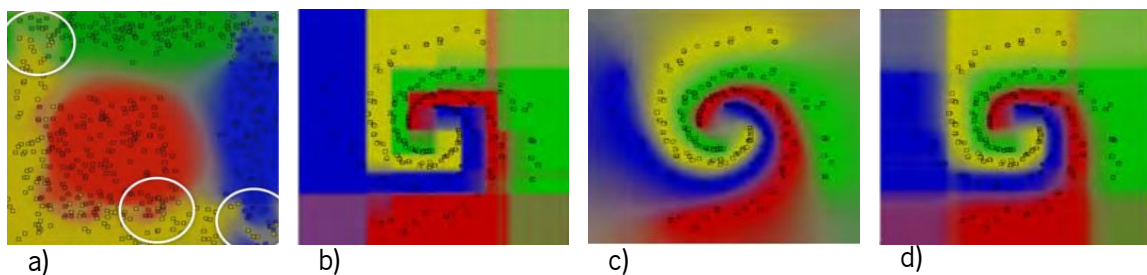


Figura 6.3 – Efeito de alguns parâmetros das *Random Forests* na divisão dos pontos de treino. Em a) demonstra-se o efeito de *overfitting*. Em b) e c) encontra-se a forma como funções lineares paralelas aos eixos e funções não lineares cónicas, respetivamente, dividem os dados de treino. Em d) observa-se o mesmo que em b), mas aumentou-se o grau de aleatoriedade. Adaptado de [71].

6.2.2.2. Propriedades de margem máxima

Nas *Support Vector Machines* a separação das classes baseia-se num hiperplano colocado de maneira a dividir os pontos, de tal forma que a distância entre o hiperplano e o ponto mais próximo de cada classe seja máxima. Nas RFs não é utilizado o mesmo princípio,

mas este pode ser replicado e utilizado como forma de estudo da eficácia do classificador para separar as classes [71].

Imagine-se o caso em que apenas existem pontos de duas classes, e que estas são linearmente separáveis, então existe uma lacuna no espaço, entre os pontos mais próximos de cada classe; a linha que se puder traçar exatamente a meio da lacuna representa a margem máxima. Cada DT, individualmente, não garante a separação dos dados segundo o princípio de margem máxima, mas a utilização de um número muito elevado de árvores numa RF, $T \rightarrow \infty$, com pouca aleatoriedade no treino, faz com que exista uma região de transição das probabilidades *a posteriori* muito suave, de tal forma que o ponto de transição entre as classes fica localizado sensivelmente no ponto de margem máxima. Isso acontece porque a linha de separação de todas as DTs fica localizada na lacuna, onde é máximo o ganho de informação (Figura 6.4(a)) [71].

Ao adicionar aleatoriedade, cada ponto individual perde importância na otimização, e não é garantido que todas as DTs dividam os dados na lacuna, ou seja, a divisão dos pontos pode não ser ótima, resultando em probabilidades *a posteriori* ainda mais suaves e no deslocamento do ponto de interseção das probabilidades das classes para um local diferente daquele de margem máxima [71]. Na Figura 6.4(b) apresenta-se o efeito do treino com maior aleatoriedade e verifica-se as probabilidades *a posteriori* no ponto de interseção dos conjuntos de pontos se afastou da linha de margem máxima, pode-se também verificar que a confiança diminuiu de modo geral, pois em nenhum local é unitária, ao contrário do que acontecia em a).

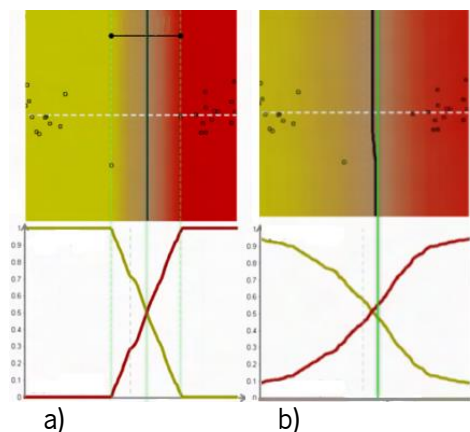


Figura 6.4 – Propriedade de margem máxima das *Random Forests*. Em a) apresenta-se a interpretação da propriedade de margem máxima em *Random Forests*, com baixo grau de aleatoriedade. Em b) o treino foi realizado com mais aleatoriedade. Adaptado de [71].

6.3. Sumário

Os classificadores são métodos de reconhecimento de padrões supervisionados, o que torna essencial a existência de dados de treino conhecidos. Estes devem ser suficientes para permitir treinar o algoritmo, e testar a sua performance, pelo que se deve adotar uma estratégia de divisão dos dados que o permita fazer, sem beneficiar nem prejudicar alguma classe.

Existem vários classificadores, sendo as DTs um deles. Estes são grafos unidirecionais em forma de árvore; os pontos transitam através dos vértices ou nós do grafo, onde alguns deles são submetidos a determinados *thresholds* sobre as suas características, que determinam para que nós devem seguir. Nas folhas existe um modelo previamente treinado, que atribui uma probabilidade a cada classe; aquela que tiver uma probabilidade maior é a classificação final.

As DTs são consideradas classificadores fracos, mas se forem combinadas várias resulta num classificador forte, as RFs. A sua principal vantagem é o facto de cada árvore ser treinada com características diferentes, e existir alguma aleatoriedade associada, o que faz com que num caso em que uma funcione mal, possa existir outra que funcione bem, resultando numa boa capacidade de generalização. Esta vantagem, aliada ao facto de ser um classificador multi-classe, foram predominantes na escolha da RF como o classificador a utilizar na segmentação de SC, SB e LCR.

Capítulo 7 Trabalho experimental

Neste capítulo descrever-se-á o sistema implementado, justificando-se as principais escolhas e alternativas. Serão, também, explicados os algoritmos utilizados, quer tenham sido implementados de raiz durante o trabalho experimental da presente dissertação, quer tenha sido utilizado uma implementação disponibilizada por outrem.

Os algoritmos especialmente implementados durante o trabalho experimental foram o detetor de EVR, utilizando o método proposto em [24], e a segmentação de SC, SB e LCR, original da presente dissertação.

O sistema implementado compreende três passos principais, o pré-processamento, a extração da região de interesse e a deteção dos EVR. Na Figura 7.1 pode-se consultar o diagrama de blocos do método desenvolvido, onde se observa a *pipeline* implementada.

Foi utilizado o pacote *Nipype* [72] para a utilização dos *softwares* de neuroimagem já existentes. O restante trabalho experimental foi implementado na linguagem de programação *Python*.

7.1. Pré-processamento

As imagens de RM utilizadas para a deteção dos EVR foram adquiridas no Hospital de Braga, tendo-se tido acesso às mesmas sem nenhum tipo de correção de artefactos, como tal, foi necessário aplicar alguns pré-processamentos. A correção da falta de homogeneidade é comum em imagens de RM, principalmente se os métodos a aplicar, posteriormente, forem muito dependentes das intensidades, como é o caso. Já a remoção do crânio ajuda a circunscrever a área de interesse ao encéfalo. Por seu lado, a normalização das intensidades toma especial importância quando se utilizam métodos supervisionados, como no caso da extração de SB, ao tornar mais consistente a gama dinâmica e o contraste, entre os diferentes sujeitos.

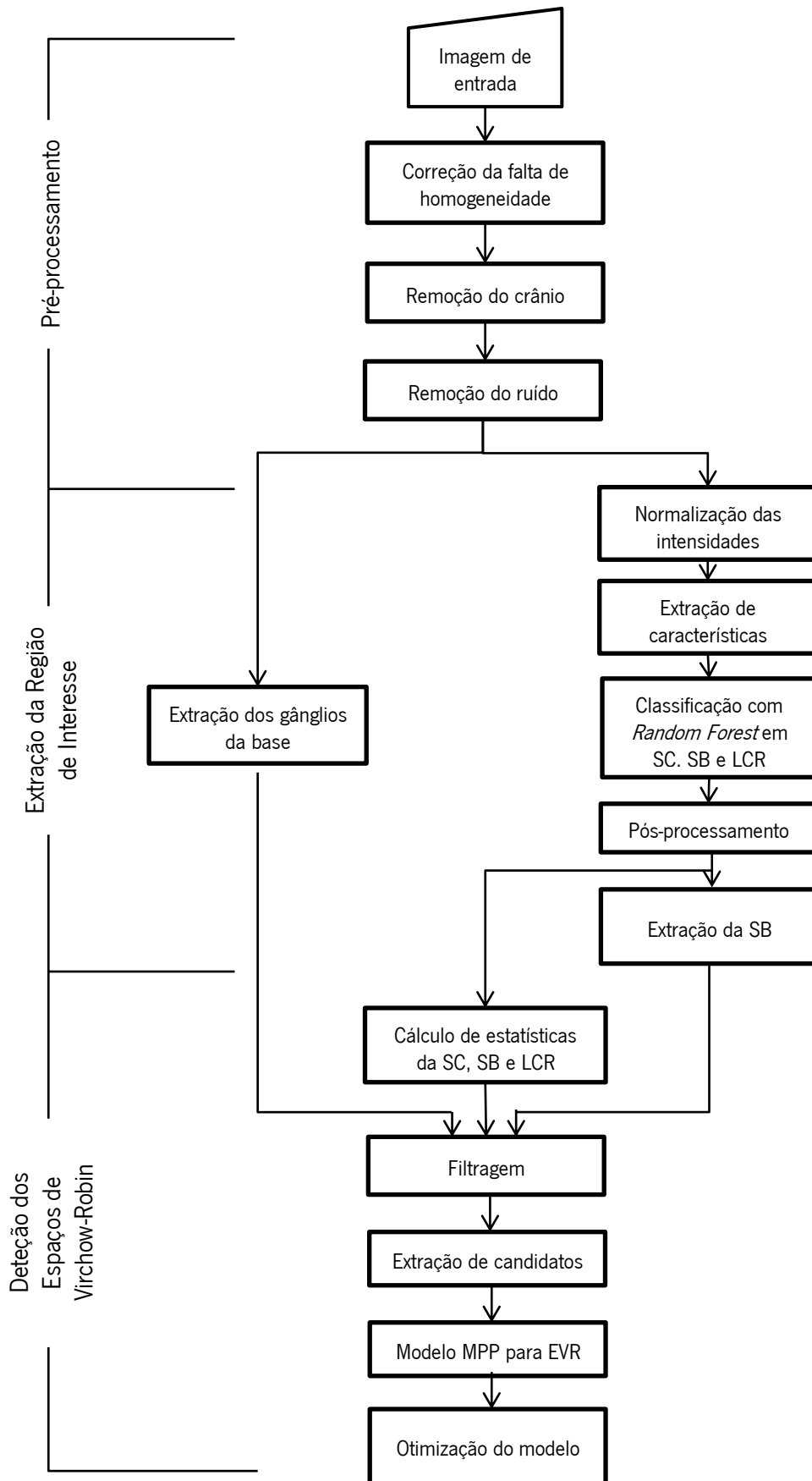


Figura 7.1 - Diagrama de blocos representativo da *pipeline* do método desenvolvido, evidenciando as etapas principais.

Outro possível pré-processamento, utilizado em [24], seria o registo para um espaço de coordenadas conhecidas com resolução isotrópica, no entanto as imagens foram adquiridas com a sequência MPAGE com resolução isotrópica de 1 mm, justificando-se, assim, a não implementação deste procedimento.

7.1.1. Correção da falta de homogeneidade

A falta de homogeneidade nas imagens de RM caracteriza-se por ser uma variação suave das intensidades, muitas vezes imperceptíveis a observadores humanos, causada por imperfeições durante a aquisição. Entre as causas apontam-se imperfeições nas bobinas de receção de radiofrequência, correntes geradas durante a aplicação de gradientes do campo magnéticos, ou mesmo a anatomia dos pacientes, quer se encontrem dentro ou fora do campo de observação [73], [74].

A correção destes defeitos implica pressupor um modelo de formação da imagem; o mais utilizado assume que a imagem adquirida, v , é o resultado da imagem verdadeira, u , sujeita a um campo de polarização, que resulta na falta de homogeneidade multiplicativa, b , e da presença de ruído proveniente do *scanner*, n , e portanto independente da aquisição. Na Equação (7.1) encontra-se a expressão matemática do modelo [73].

$$v(\mathbf{x}) = u(\mathbf{x})b(\mathbf{x}) + n(\mathbf{x}) \quad (7.1)$$

Pode-se fazer a correção durante a aquisição da imagem, através da calibração do *scanner* ou da melhoria dos processos, ou depois. Os métodos que utilizam a primeira abordagem denominam-se prospetivos, enquanto os que utilizam a segunda são chamados retrospectivos [73].

Uma vez que apenas se teve acesso às imagens, apenas se considerou os métodos retrospectivos, apesar de estes poderem remover algumas variações suaves nas intensidades dos tecidos [74]. Estes podem ser divididos em métodos de filtragem, adaptação de superfície, segmentação e baseados em histogramas. Os primeiros pressupõem que a variação das intensidades é suave, então uma filtragem passa-alto deveria conseguir recuperar a imagem original. O segundo tipo de correções adapta uma superfície paramétrica a um conjunto de características da imagem, por exemplo as intensidades ou o gradiente, e a superfície resultante corresponde à estimativa da falta de homogeneidade. Também é possível fazer as correções durante a segmentação, alternando entre os dois processos. Por fim, o último tipo de abordagem retrospectiva, tal como o nome indica, consegue estimar a falta de homogeneidade com base no histograma das imagens [73].

Pretendia-se utilizar um método já implementado, existindo, pelo menos, duas alternativas: o método *FMRIB's Automated Segmentation Tool* (FAST) [44] incluído no pacote *FMRIB Software Library* (FSL) [75], o método *Nonparametric Nonuniform Intensity Normalization* (N3) [74] incluído no *Freesurfer* e uma sua variante denominado *Nick's Nonuniform Intensity Normalization* (N4) [76].

O método FAST é um algoritmo de segmentação, que corrige a falta de homogeneidade em simultâneo, retornando, também, a imagem original corrigida, enquanto os métodos N3 e N4 baseiam-se no histograma. O FAST exige que o utilizador defina, *a priori*, o número de classes que devem ser segmentadas, porém as imagens de RM utilizadas foram adquiridas de pessoas idosas que sofreram AVCs, pelo que a decisão do número de classes não é trivial devido às zonas danificadas, para além disso é necessário que o crânio tenha sido extraído previamente. Por outro lado, os métodos baseados em histogramas são totalmente automáticos, são independentes da sequência de RM, podem ser aplicados a imagens com patologias, não necessitam que se especifique um modelo para as intensidades dos tecidos e podem ser aplicados com a presença do crânio, o que pode ser benéfico para a extração do mesmo numa etapa posterior [73], [74], [76]. Portanto, optou-se pela utilização do método N3 ou N4.

7.1.1.1. Método N3 e N4

O método N3 pressupõe que a falta de homogeneidade esbate a imagem, devido à sua baixa frequência, por isso o objetivo é descobrir o campo de baixas frequências para que se maximize a frequência da imagem [74].

Considerando o modelo de formação da imagem da Equação (7.1), os autores pressupõe que a mesma é livre de ruído; e fazendo o logaritmo torna a falta de homogeneidade aditiva, transformando o modelo na Equação (7.2). Em seguida, se V , U e B forem as distribuições de probabilidades de \hat{v} , \hat{u} e \hat{b} , respetivamente, a distribuição da imagem \hat{v} resulta da convolução de U com B [74].

$$\hat{v}(\mathbf{x}) = \hat{u}(\mathbf{x}) + \hat{b}(\mathbf{x}) \quad (7.2)$$

A distribuição de probabilidades V é calculada a partir do histograma da imagem, B é aproximada a partir de uma distribuição gaussiana e U deve ser estimada. Esta estimativa é calculada através da deconvolução entre V e uma gaussiana estreita, que representa B . Assim, pode-se estimar \hat{u} ($E[\hat{u}|\hat{v}]$) que é subtraída à imagem corrompida, \hat{v} , resultando num mapeamento entre as intensidades da imagem e o campo da falta de homogeneidade, \hat{b}_s

(Equação (7.3)). Porém, este mapeamento deve ser suavizado com um operador S , tendo os autores optado por usar *B-splines* (Equação (7.3)) [74].

$$\hat{b}_s(\hat{v}) = S\{\hat{v} - E[\hat{u}|\hat{v}]\} \quad (7.3)$$

Este é um processo iterativo, que em cada iteração a imagem corrigida da iteração anterior é utilizada para estimar o campo de falta de homogeneidade \hat{b}_s a partir da imagem inicial, e estima novamente a imagem corrigida, tal como se pode perceber pelas equações (7.4) que representam o processo. O processo repete as estimativas até que não se verifiquem mudanças em \hat{b} ou U , resultantes de deconvolução de gaussianas estreitas a V , ou seja, em cada iteração é estimada a falta de homogeneidade global. Este processo iterativo corresponde a uma procura do campo de não homogeneidades, no sentido de campos com gaussianas cada vez mais largas [74].

$$\begin{aligned} \hat{u}^1 &= \hat{v} - S\{\hat{v} - E[\hat{u}|\hat{v}]\} \\ \hat{u}^2 &= \hat{v} - S\{\hat{v} - E[\hat{u}|\hat{u}^1]\} \\ &\vdots \\ \hat{u}^n &= \hat{v} - S\{\hat{v} - E[\hat{u}|\hat{u}^{n-1}]\} \end{aligned} \quad (7.4)$$

Nos casos em que o campo magnético é mais forte (3T), as variações de intensidade da falta de homogeneidade são maiores. Assim, o método N3 tem o parâmetro da distância entre os nós da malha da *spline* que pode ser reduzido para compensar a maior frequência, até 30 mm, sendo o seu valor padrão de 200 mm. No entanto, se a redução for muito acentuada, o método falha [76].

Assim sendo, o método N4 propõe-se melhorar o anterior, através da utilização de uma *B-spline* mais rápida e robusta, mas também utilizando um método de otimização mais eficiente [76].

A abordagem com uma nova *B-spline* consegue utilizar distâncias mais pequenas entre os nós da malha de forma mais estável, e é mais eficiente, computacionalmente. Além destas vantagens, a malha é inicializada com uma determinada distância, e vai sendo reduzida ao longo das iterações, ou seja, vai aumentando a resolução, o que não acontecia com o N3, cuja distância era fixa. Esta nova abordagem é mais robusta à presença de ruído, para além de tornar o método N4 mais fiável para campos mais fortes [76].

Também o processo iterativo para estimar o campo de falta de homogeneidade foi alterado, tornando a otimização mais eficiente. Assim, ao contrário do método N3, a falta de homogeneidade é sempre estimada relativamente à imagem corrigida na iteração anterior, ou

seja, vai-se convergindo para a verdadeira solução, ao contrário de se estimar sempre a falta de homogeneidade total. Esta abordagem demonstrou convergência superior [76]. Nas Equações (7.5) encontra-se o processo iterativo do N4, e pode-se perceber melhor a diferença para o N3.

$$\begin{aligned}
 \hat{u}^1 &= \hat{v} - S^*\{\hat{v} - E[\hat{u}|\hat{v}]\} \\
 \hat{u}^2 &= \hat{u}^1 - S^*\{\hat{u}^1 - E[\hat{u}|\hat{u}^1]\} \\
 &\vdots \\
 \hat{u}^n &= \hat{u}^{n-1} - S^*\{\hat{u}^{n-1} - E[\hat{u}|\hat{u}^{n-1}]\}
 \end{aligned}
 \tag{7.5}$$

Na validação do N4, este provou ser superior do que o N3 à medida que a distância das *splines* ia diminuindo, mas também na presença de ruído.

Durante o processo experimental, foi testado se a aplicação de um ou outro algoritmo era relevante durante a extração do encéfalo. Para isso, aplicou-se o método N3 e N4 às imagens disponíveis, e verificou-se que, de facto, corrigindo a falta de homogeneidade com N4 conduzia a remoções do crânio mais eficazes, tal como se pode observar na Figura 7.2. Assim sendo, optou-se por utilizar o método N4, distribuído no pacote ANTs [77].

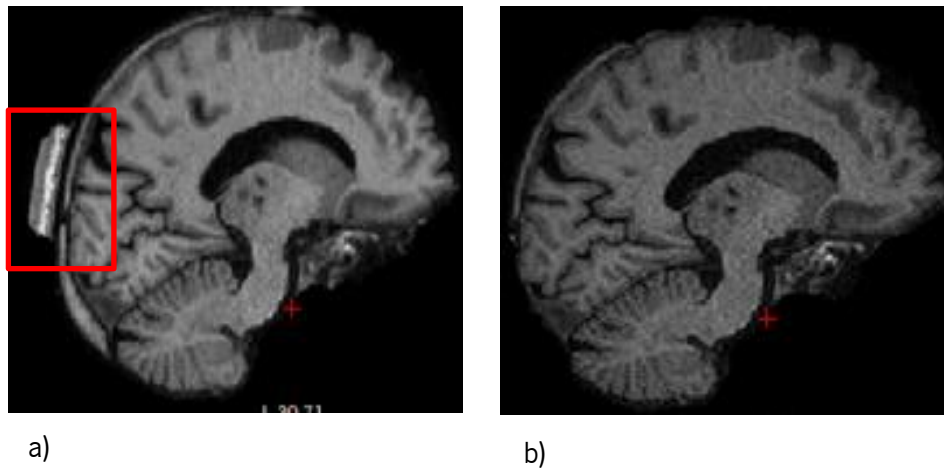


Figura 7.2 – Efeito da correção da falta de homogeneidade com N3 e N4. Em a) foi utilizado o método N3, enquanto em b) foi o N4. Aplicando depois o algoritmo de remoção do crânio, verifica-se que na imagem corrigida com N3, ainda fica uma porção de crânio por remover.

Tal como na validação realizada em [76], optou-se por inicializar a distância das *B-splines* com o valor 200 mm, por ser difícil avaliar o seu efeito. Na Figura 7.3 observa-se uma imagem de RM do encéfalo, a mesma imagem corrigida e a estimativa da falta de homogeneidade, e pode-se verificar que a intensidade dos tecidos não é uniforme, particularmente se se comparar o centro com a periferia, apesar de a falta de homogeneidade ser pouco acentuada na imagem original, o que é positivo.

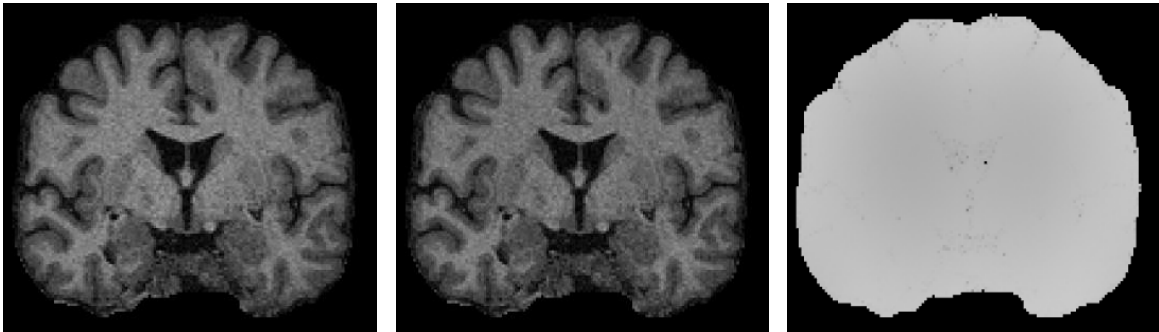


Figura 7.3 - Correção da falta de homogeneidade com o método N4. Da esquerda para a direita, observa-se a imagem original, a imagem corrigida e a estimativa da falta de homogeneidade.

7.1.2. Remoção do crânio

A remoção do crânio é importante num processo de segmentação de tecidos do encéfalo. Neste sentido, pretende-se excluir da imagem os tecidos que não são relevantes, o que contribui para que a segmentação seja mais precisa, pois retiram-se possíveis falsos positivos, e restringe-se o problema a uma região de interesse mais pequena, o que também melhora a performance do sistema. Assim, neste tipo de procedimento pretende-se que, no final, apenas esteja presente o encéfalo, e LCR extracerebral.

Para atingir este objetivo, procurou-se métodos de remoção do crânio já implementados e disponíveis para utilização, tendo-se como opções o *Brain Extraction Tool* (BET) [78], do pacote de *software* FSL [75], e o método utilizado pelo *Freesurfer*, que se irá denominar por abordagem híbrida [79]. O primeiro utiliza um modelo deformável, que evolui com base em forças para se adaptar à superfície do encéfalo [78]; por outro lado, o segundo segue uma abordagem híbrida com *watershed* e modelos deformáveis [79].

Foram testados ambos os métodos, tendo-se percebido que o BET era dependente de um parâmetro de *threshold*, que em poucos casos foi igual para diferentes sujeitos, sendo necessário fazer ajustes. Além disso, este parâmetro mostrou ser sensível, uma vez que pequenas diferenças correspondiam a remoção de grandes porções de SC e SB. Por outro lado, a abordagem híbrida tinha poucos parâmetros, apenas se escolheu que deveria utilizar um atlas, para tornar a remoção mais robusta, deixando-se as restantes opções por defeito. Na Figura 7.4 pode-se observar dois planos axiais do mesmo encéfalo, mas em a) removeu-se o crânio com a abordagem híbrida, enquanto em b) foi com o BET, e como se pode perceber, há regiões de tecido do encéfalo que o BET removeu, erradamente. Assim sendo, como se pretendia um sistema automático, foi escolhida a abordagem híbrida para extrair o encéfalo, e verificou-se que

na grande maioria dos casos removeu bem o crânio, existindo casos em que removeu um pouco mais do que devia, mas foi aceitável, tendo em conta que nunca foi necessário ajustar nenhum parâmetro.

Além disso, os autores da abordagem híbrida, em [79], testaram a sua proposta contra outros métodos de remoção do crânio, incluindo o BET. Foram utilizados dois conjuntos de imagens, contendo sujeitos jovens, idosos, dementes e com Doença de Alzheimer, sendo que a abordagem híbrida conseguiu 0.885 e 0.858, em Jaccard, para cada conjunto, respetivamente, contra 0.483 e 0.570 do BET, reforçando a escolha da abordagem híbrida.

A abordagem híbrida só se aplica à sequência T1. Portanto, quando se se utilizou outras sequências, primeiro removeu-se o crânio de T1, uma vez que se tinha sempre esta sequência, e depois criou-se uma máscara, com base na imagem que apenas continha o encéfalo, e aplicava-se às restantes sequências, desde que estivessem todas registadas, o que era verdade.

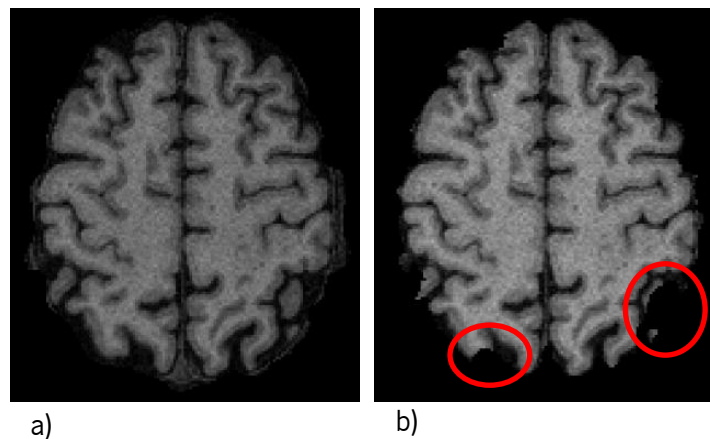


Figura 7.4 – Remoção do crânio. Em a) observa-se a remoção do crânio com o algoritmo presente no *Freesurfer*, enquanto em b) foi utilizado o BET. Os círculos a vermelho realçam espaços onde foi removido tecido do encéfalo.

7.1.3. Normalização das intensidades

As intensidades das imagens de RM não são constantes entre aquisições, mesmo utilizando protocolos iguais, com a mesma pessoa e o mesmo *scanner*. Este facto torna-se relevante para processos de segmentação, particularmente se implicar uma fase de aprendizagem; por exemplo, se for utilizado um classificador supervisionado [80].

Zikic et al, no seu trabalho de segmentação de tumores cerebrais, alinha a média de todas as imagens através de um fator de multiplicação global. Contudo, esta abordagem não tem em consideração a distribuição das intensidades, o que pode resultar no alinhamento entre *voxels* com significados diferentes [10].

Por outro lado, *Nyul et al.* propõem um método de normalização das intensidades de imagens de RM através da modificação do histograma. Neste caso, utiliza-se um histograma de referência, e as intensidades de todas as outras imagens são modificadas para que os seus histogramas se transformem numa distribuição semelhante à referência. Desta forma, procura-se que as regiões do histograma que representem os mesmos tecidos fiquem alinhadas, com o objetivo de tornar as intensidades e o contraste mais consistente entre os sujeitos [80].

Assim sendo, optou-se por utilizar a normalização de *Nyul et al.*, que se encontra implementada no pacote de *software Insight Toolkit* [81]. As sequências devem ser normalizadas independentemente, isto é, para normalizar T1 deve-se escolher uma referência T1, enquanto para FLAIR deve ser escolhida uma referência FLAIR.

7.1.3.1. Transformação do histograma

A forma de transformação do histograma apresentada em [80] pressupõe a existência de um histograma de referência, e um conjunto de marcas que servirão para fazer o emparelhamento entre as distribuições.

Começa-se, então, por calcular a intensidade mínima e máxima do histograma, m_1 e m_2 , respetivamente, e por se excluir as caudas do mesmo, uma vez que geralmente correspondem a *outliers*. As intensidades a excluir são aquelas inferiores ao primeiro percentil, pc_1 , e superiores ao último percentil, pc_2 , e o intervalo entre estes dois percentis é designado por intensidade de interesse. Em seguida deve-se selecionar as marcas do histograma, μ_p , onde p representa o percentil. Na Equação (7.6) pode-se observar alguns conjuntos exemplo de marcas [80], enquanto na Figura 7.5(a) se encontra um exemplo de colocação das marcas num histograma.

$$\begin{aligned} L_1 &= \{pc_1, pc_2, \mu_{50}\} \\ L_2 &= \{pc_1, pc_2, \mu_{25}, \mu_{50}, \mu_{75}\} \\ L_3 &= \{pc_1, pc_2, \mu_{10}, \mu_{20}, \dots, \mu_{90}\} \end{aligned} \quad (7.6)$$

Depois de determinados os parâmetros anteriores para o histograma de referência obtém-se uma escala de referência, delimitada pelo intervalo $[s_1 = pc_1, s_2 = pc_2]$. Então, para normalizar as intensidades de uma nova imagem, deve-se determinar as marcas no seu histograma, e mapeá-las linearmente para a escala de referência, no intervalo $[s_1, s_2]$. Em seguida, faz-se um novo mapeamento linear, desta vez entre marcas. Assim sendo, se o índice i representar marcas da imagem que se pretende normalizar, o índice k designar marcas do histograma de referência, de tal forma que i e k tenham a mesma gama de valores, ou seja, o

número de marcas na referência e na imagem a normalizar deve ser o mesmo, e existirem l marcas μ , então começa-se por mapear $[pc_{1i}, \mu_{1i}]$ para $[s_{1k}, \mu_{1k}]$, depois $[\mu_{1i}, \mu_{2i}]$ para $[\mu_{1k}, \mu_{2k}]$, e assim sucessivamente até $[\mu_{li}, pc_{2i}]$ para $[\mu_{lk}, s_{2k}]$. Na Equação (7.7) encontra-se a expressão para o mapeamento linear das intensidades de $[pc_{1i}, pc_{2i}]$ para $[s_1, s_2]$, onde x é a intensidade original e x' é a intensidade mapeada; a mesma equação pode ser utilizada para o mapeamento entre marcas, fazendo as necessárias alterações nas variáveis, para essas marcas [80], [82]. Já na Figura 7.5(b) pode-se observar, graficamente, o processo de normalização linear entre marcas.

$$x' = s_1 + \frac{x - pc_{1i}}{pc_{2i} - pc_{1i}} (s_2 - s_1) \quad (7.7)$$

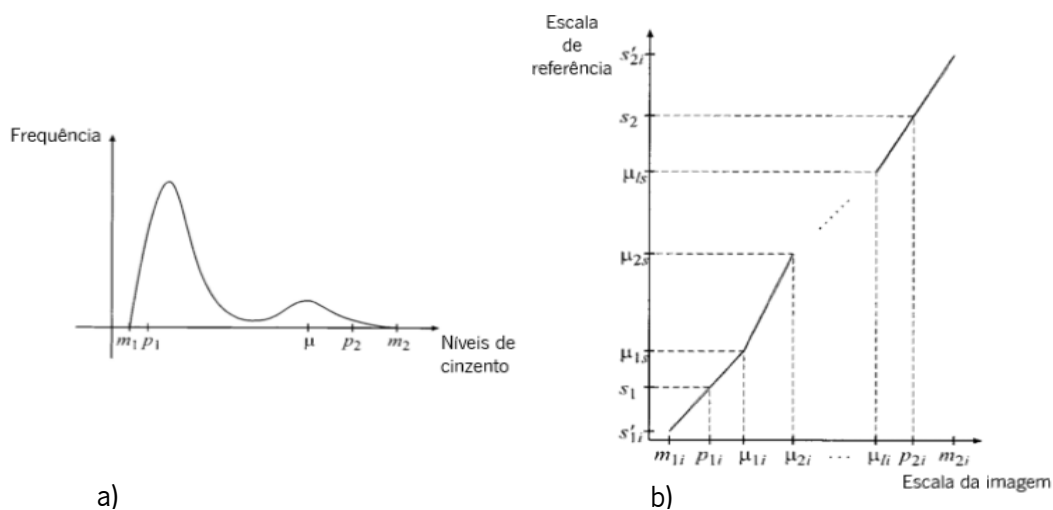


Figura 7.5 – Normalização das intensidades. a) Exemplo de escolha das marcas numa distribuição bimodal, na qual se optou por colocar a marca μ na segunda moda. b) demonstração do mapeamento linear entre o histograma da imagem que se pretende normalizar e o histograma de referência. Adaptado de [80].

Deve-se ter o cuidado de se excluir os *voxels* de fundo, durante a construção dos histogramas [80].

Após análise do algoritmo, pode-se apontar que o local de colocação das marcas é crítico, uma vez que se não for o correto, e as distribuições forem diferentes, pode resultar na compactação de algumas gamas de intensidades, ou então no alongamento de outras gamas, de forma indevida.

Como já foi referido, foi utilizada a implementação constante no *Insight Toolkit*, tendo-se selecionado duas marcas para os histogramas, e que o algoritmo deveria excluir o fundo. O número de *bins* utilizado para a construção dos histogramas foi aquele que estava definido por

defeito, ou seja, 1024. O número de marcas foi selecionado através da visualização das imagens.

7.1.4. Remoção de ruído

Na aquisição de imagens de RM existe um compromisso entre a resolução e a relação sinal-ruído. As imagens utilizadas para a detecção de EVR possuíam resolução isotrópica de 1 mm, por isso o ruído era relevante.

Na RM o ruído segue a distribuição Rician [83]. *Diaz et al*, em [83], realizaram um estudo comparativo do efeito da remoção de ruído, na segmentação de tumores cerebrais, em imagens de RM. Após os vários testes, concluíram que todos os métodos testados eram eficazes naquilo a que se propunham, mas não melhoravam o contraste, sendo a melhor performance atribuída ao *Non-local Means Filter* com distribuição Rician proposto por Manjón, em [84]. Em [85] comprova-se que a utilização de um *Non-local Means Filter* Riciano melhora a qualidade da segmentação de SB, SC e LCR em imagens com elevado ruído. Por estes motivos, optou-se por utilizar um método baseado neste tipo de filtros, proposto por *Manjón et al*, em [86], denominado *Prefiltered Rotationally Invariant Non-Local Mean (PRINLM) Filter* cuja implementação foi disponibilizada no *website* de um dos autores².

7.1.4.1. Prefiltered Rotationally Invariant Non-Local Mean Filter

Pode-se filtrar o ruído de uma imagem de RM calculando a sua transformada discreta do cosseno, para cada bloco 3D (4x4x4). Em seguida, aplica-se um *threshold* igual a $2,7\sigma$, onde σ é a estimativa do desvio padrão do ruído, para se obter uma primeira estimativa da imagem filtrada [86].

Por outro lado, a ideia do *Non-Local Mean Filter* é que numa imagem existe informação redundante, isto é, há muitas regiões semelhantes. As regiões denominam-se *patches*, sendo constituídas pelo *voxel* sob análise (central) e a sua vizinhança 3D. O processo de remoção de ruído consiste na substituição do *voxel* pela média das intensidades dos *voxels* centrais dos *patches* que são parecidos entre si. No entanto, podem existir *patches* semelhantes, mas rodados, sendo necessário utilizar medidas de rotação para os descobrir, passando a denominar-se *Rotational Invariant Non-Local Mean Filter*. Este novo filtro é poderoso, pois restaura a imagem sem afetar muito a sua estrutura, mas é sensível à presença de muito ruído para estimar os *patches* mais idênticos [86].

² <https://sites.google.com/site/pierrickcoupe/software/denoising-for-medical-imaging/mridenoising>, em 20/10/2013

O filtro PRINLM começa por filtrar a imagem utilizando o método da transformada discreta do cosseno. Posteriormente utiliza essa imagem pré-filtrada para identificar os *patches* que são semelhantes, de forma invariante à rotação, e utiliza-os para remover o ruído da imagem original com a abordagem *Non-Local Means*. Assim, consegue-se tirar partido do *Rotational Invariant Non-Local Mean Filter*, mesmo em imagens com muito ruído. A distribuição Rician para o ruído é tida em consideração quer no cálculo da semelhança dos *patches*, quer na estimativa do ruído no método da transformada discreta do cosseno. Os autores reportam os melhores resultados, entre os algoritmos comparados, que referem ser o estado da arte [86].

Relativamente aos parâmetros, definiu-se que deveria utilizar o modelo de ruído Riciano, que a dimensão do *patch* seria $3 \times 3 \times 3$ e deveria fazer a procura de *patches* semelhantes no espaço $7 \times 7 \times 7$ que envolve o *voxels* que está a ser analisado. A determinação dos parâmetros foi feita visualmente, e verificou-se que aumentar a dimensão do *patch*, ou da região de procura resultava em imagens indistinguíveis, optando-se pelas referidas dimensões, em prol da velocidade de processamento, sendo as definições por defeito. Na Figura 7.6 encontra-se o resultado da remoção de ruído com o filtro PRINLM, e verifica-se que é bastante eficaz. Na imagem ampliada nota-se que as pequenas estruturas que podem ser lesão ou EVR são preservadas, embora aquelas que tivessem intensidade um pouco mais elevada possam ter ficado mais esbatidas.

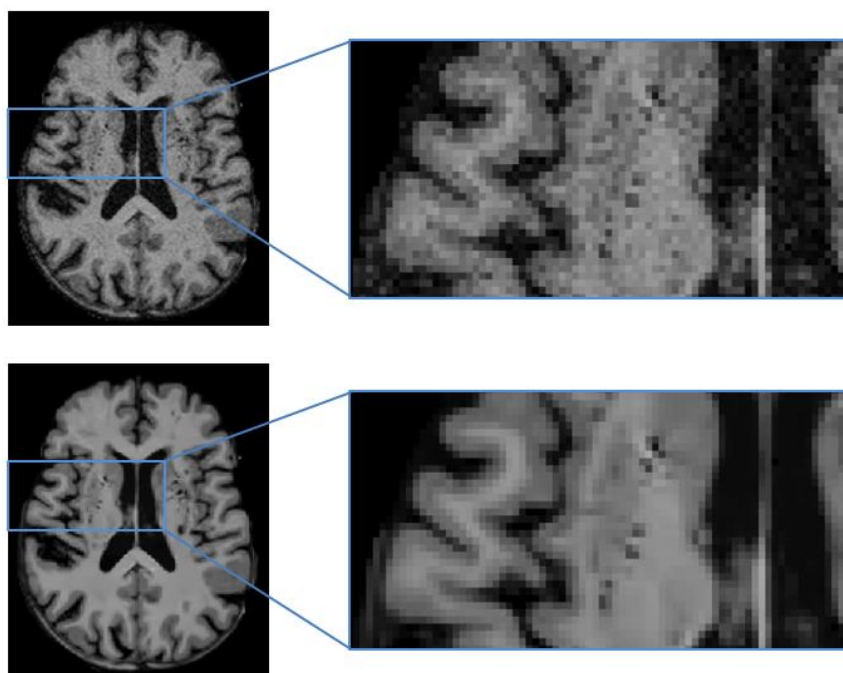


Figura 7.6 - Efeito da remoção de ruído, num plano axial. Em cima observa-se a imagem de RM com ruído, enquanto em baixo se encontra a mesma imagem filtrada.

7.2. Extração da Região de Interesse

Tal como já foi referido no Capítulo 2, os estudos que relacionam os EVR com determinadas patologias, na sua maioria, fazem a contagem do número dessas estruturas em determinadas regiões, nomeadamente nos gânglios da base, região subinsular e centro semioval (ou SB cerebral) [3], [4], [30], [35].

Assim sendo, deve-se extrair as regiões de interesse antes de se aplicar o algoritmo de deteção dos EVR. No procedimento para segmentação subcortical extrai-se os gânglios da base, e procura-se, também, incluir a região subinsular, enquanto num outro procedimento extrai-se a SB cerebral.

7.2.1. Segmentação subcortical

Os gânglios da base são o conjunto do putamen, globo pálido e núcleo caudado, sendo que alguns autores ainda incluem a substância negra e núcleo subtalâmico. Por sua vez, a região subinsular encontra-se entre o putamen e a ínsula (ver Figura 2.3 e Figura 2.4).

Utilizou-se o *Freesurfer* para segmentar a região subcortical, uma vez que que o seu procedimento consegue segmentar até 37 estruturas, e o registo utilizado é robusto à variabilidade anatómica típica em cérebros de pessoas idosas, sendo este o grupo etário alvo da dissertação [87]. Para além disso, o conjunto de dados utilizados para a deteção de EVR foi adquirido com a sequência MPRAGE, cujo protocolo foi desenvolvido paralelamente com o *Freesurfer*, para que exista bom contraste entre a SB, SC e LCR [88], reforçando a seleção do algoritmo de segmentação subcortical.

O algoritmo de segmentação do *Freesurfer* utiliza uma abordagem estatística, ao considerar que as intensidades de cada região seguem uma distribuição gaussiana, ao aplicar atlas de probabilidades (de forma a incorporar informação do espaço) e ao utilizar um modelo MRF [87], [89].

O método regista o atlas na imagem, resultando numa transformação L , que faz a correspondência entre as coordenadas dos dois espaços. Assim sendo, pretende-se maximizar a Equação (7.8), que surge da regra de Bayes [87].

$$p(W|I, L) \propto p(I|W, L)p(W) \quad (7.8)$$

onde W representa a classe, I as intensidades, $p(W|I, L)$ a probabilidade *a posteriori* de um dado *voxel* com intensidade I pertencer à classe W , $p(I|W, L)$ é a probabilidade condicionada

do *voxel* modelizada com a distribuição gaussiana e $p(W)$ é a probabilidade *a priori* da classe W [87].

O atlas permite que em cada região do encéfalo apenas se considerem algumas probabilidades condicionadas $p(W|I, L)$ e classes, uma vez que se sabe qual é o número de possíveis estruturas aí presentes [87], [89].

O modelo MRF é utilizado na probabilidade *a priori*, sendo anisotrópico e não estacionário, isto é, o peso dado aos *voxels* da vizinhança não é igual para todos os cliques e varia em função da localização espacial, e das estruturas mais prováveis de estarem aí presentes [87], [89].

Em [89] melhora-se o método de [87], ao fazer-se a normalização das intensidades de cada estrutura do atlas para as intensidades do sujeito a ser processado, através da transformação linear dos histogramas, melhorando os resultados, e permitindo, até, utilizar-se protocolos de aquisição e *scanners* diferentes daqueles que foram utilizados para adquirir os sujeitos de treino.

O resultado da segmentação subcortical do *Freesurfer* é uma imagem com diversas *labels*. O *software* altera a imagem de RM para as dimensões 256X256X256, pelo que no final deve ser convertida para a dimensão original, utilizando a função *mri_label2vol* incluída no *Freesurfer*. Extraíu-se, então, as estruturas marcadas com os números 11, 12, 13, 51, 52 e 53, correspondentes ao núcleo caudado, putamen e globo pálido, em cada hemisfério. Extraíu-se, ainda, a classe 16, sendo esta o tronco cerebral, de forma a incluir a substância negra, porque não existia nenhuma *label* dedicada, para além disso, o tipo III de localizações de EVR refere-se ao mesencéfalo. Por fim, transformou-se a imagem com as várias regiões de interesse numa imagem binária, criando-se uma máscara.

Como os gânglios da base são um dos principais locais de AVC, *microbleeds* e EVR podem existir regiões hipointensas em T1 erroneamente classificadas, que depois poderiam não ser utilizadas no algoritmo de deteção de EVR. Assim, aplicou-se um pós-processamento sobre a máscara das regiões de interesse, correspondendo à utilização de uma função para fechar buracos, segundo as três dimensões do espaço.

Em seguida, fez-se duas dilatações morfológicas da máscara com um objeto estruturante 3D, em cruz. Assim, consegue-se incluir, pelo menos parcialmente, a região subinsular. Este procedimento faz com que se incorpore alguma SB cerebral na região de interesse, mas considerou-se que seria mais relevante adicionar a referida região. Além disso,

como o procedimento seria igual para todos os sujeitos, a possível polarização introduzida seria constante e conhecida. Finalmente, também se adicionou na região de interesse os conjuntos com menos de 50 *voxels* que possam não ter sido fechados e sejam consequência de algum AVC.

Na Figura 7.7 pode-se observar a região de interesse extraída, antes e após o pós-processamento da segmentação do *Freesurfer*. Pode-se concluir que o pós-processamento obteve o efeito desejado, ao incluir regiões que não foram segmentadas, mas que aparentemente o deveriam ter sido. Também se confirma a inclusão de alguma SB, mas visualmente aceita-se, dada a melhoria da região de interesse em geral.

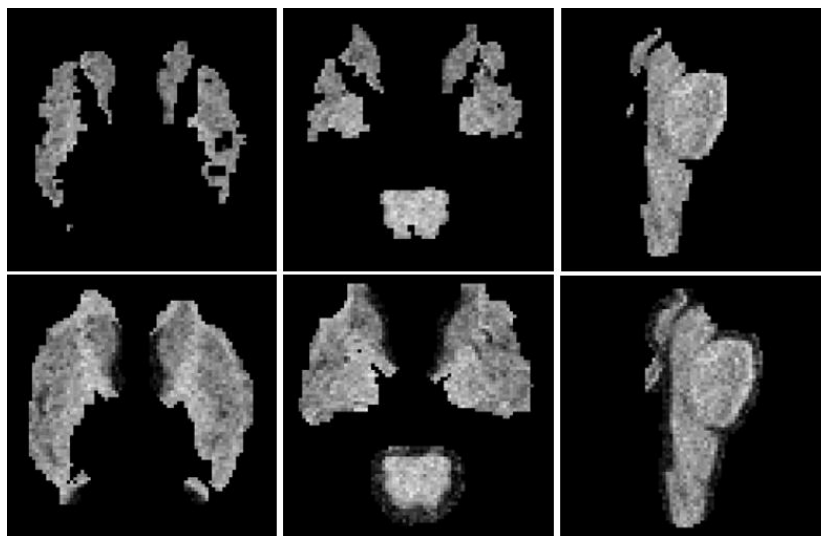


Figura 7.7 - Extração dos gânglios da base e região subinsular. Na primeira linha observa-se a região de interesse extraída apenas com a segmentação do *Freesurfer*, enquanto na segunda linha se observa o pós-processamento da segmentação. Da esquerda para a direita encontram-se os planos axial, coronal e sagital da imagem de RM, para o mesmo sujeito e mesmos *slices*.

7.2.2. Segmentação de tecidos do encéfalo e extração da substância branca

Tal como já foi referido, alguns estudos que se debruçam sobre a presença de EVR restringem a sua contagem a uma outra região de interesse, a SB cerebral [3], [30]. No entanto, a sua extração não é trivial, uma vez que os dados foram adquiridos a partir de sujeitos que sofreram AVCs, podendo este acontecer na SB.

As soluções disponíveis em pacotes de *software* são maioritariamente abordagens estatísticas, que utilizam misturas de gaussianas e MRF para segmentar, como o FAST [44] ou o *Atropos* [90] do pacote ANTs [77]. O problema dos referidos métodos é o facto de o número de classes a segmentar ter que ser definido antes da aplicação e terem que estimar os parâmetros com base apenas na imagem em análise, repercutindo-se na performance da segmentação, pois

nem todas as imagens de RM têm lesões, ou outras anormalidades que impliquem mais categorias do que as dos tecidos normais. Por exemplo, se se definir menos classes do que se deveria, pode haver fusão de tecidos, resultando no cálculo deficiente dos parâmetros dos modelos; se por outro lado, o número de classes for superior, pode-se perder o significado de algumas delas, nomeadamente se os AVCs forem de muito reduzida dimensão, de tal forma que não seja possível calcular as suas estatísticas com confiança, o que pode resultar na divisão de uma classe em vários *clusters*.

Pelos motivos supramencionados, optou-se por projetar e implementar um algoritmo para a segmentação de SC, SB e LCR, que fosse robusto à presença de lesões, ou outras anormalidades dos tecidos. As abordagens referidas anteriormente são não-supervisionadas, e apesar de haver alguns métodos desta categoria que não necessitam do número de classes como *input*, como o *mean shift*, levantam o problema da fusão de classes no final, já que muitas vezes resultam em múltiplos *clusters*. Mas mais importante, escolher o modelo a ser utilizado para a agregação dos dados não é trivial, bem como a otimização correta dos parâmetros.

Assim sendo, uma abordagem supervisionada que conseguisse aprender o modelo a partir dos dados, capaz de generalizar mesmo com a presença de intensidades anormais dos tecidos, foi considerada a mais adequada. Optou-se, então, pela utilização de RFs devido à sua boa capacidade de generalização para novos dados (pois os problemas que afetam o encéfalo são variados e podem ocorrer em diversas localizações), capacidade de paralelização, e serem, naturalmente, classificadores multi-classe.

Na Figura 7.8 encontra-se a *pipeline* do algoritmo de segmentação do LCR e tecidos SB e SC. Os procedimentos incluídos na etapa de pré-processamento foram descritos nas secções anteriores. Contudo, o método de remoção do crânio deve ser utilizado somente em T1. Portanto, para extrair o encéfalo noutras sequências, fez-se os registo delas para T1; posteriormente criou-se uma máscara do encéfalo de T1, e aplicou-se nas restantes sequências.

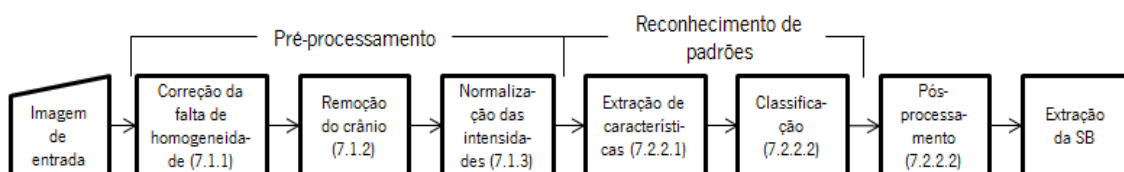


Figura 7.8 - *Pipeline* do algoritmo de segmentação de tecidos, e extração da substância branca.

Os dados existentes para treinar o classificador consistiam em três sequências de RM (T1w, T1w IR e FLAIR) registadas em T1, para cinco sujeitos; adquiridas num *scanner* de 3T,

com dimensão dos *voxels* de 0,958 mm x 0,958 mm x 3 mm. Os sujeitos tinham todos idade superior a 50 anos, e apresentavam vários graus de atrofia e lesões da SB. Existiam duas marcações, uma marcava SC cortical, gânglios da base, SB, lesões da SB, LCR no espaço extracerebral, ventrículos, cerebelo e tronco cerebral. A outra marcava SC, SB e LCR, fundindo as classes de SC cortical e gânglios da base, SB e lesões da SB e LCR do espaço extracerebral e ventrículos, e removendo o cerebelo e o tronco cerebral. Foi utilizada a segunda marcação manual para treinar o algoritmo. Este conjunto de dados foi obtido através do *MICCAI Grand Challenge on MR Brain Image Segmentation 2013*.

7.2.2.1. Extração de características

Foi escolhido e testado um conjunto de características relevantes: as intensidades, a informação da vizinhança, as probabilidades *a posteriori* de cada classe, o atlas de probabilidades de tecidos e o módulo do gradiente. As características foram extraídas no plano axial, tendo-se utilizado vizinhanças, quando necessário, bidimensionais porque o método foi treinado e validado num conjunto de imagens cuja distância entre *slices* axiais media 3 mm.

Intensidades – utilizou-se as três sequências existentes, e a diferença de intensidades entra cada par de sequências, para fazer sobressair as particularidades de cada uma (Figura 7.9(a-c)), resultando três intensidades e três diferenças.

Informação da vizinhança – para cada *voxel* extraiu-se quatro vizinhanças 2D no plano axial, correspondendo a quadrados cujas arestas mediam 3 mm, 9 mm, 15 mm e 19 mm, respetivamente. Para cada vizinhança e sequência calculou-se três medidas que codificam as intensidades da região, a média, soma e mediana, utilizando funções implementadas nos pacotes *numpy* e *scipy*. Utilizou-se, ainda, a diferença das medidas entre cada par de sequências, resultando, no total, em 72 características.

Probabilidades *a posteriori* – estas probabilidades são utilizadas como característica, mas resultam de uma segmentação prévia da imagem com o pacote *Atropos* [90], que disponibiliza as probabilidades *a posteriori* dos tecidos como *output*. Assim, a RF vai ter em consideração uma classificação já existente da imagem, estando-se perante uma cascata de classificadores. Portanto, pretende-se melhorar a classificação do método aplicado pelo *Atropos*, e contornar as suas limitações, nomeadamente o seu comportamento na presença de lesões.

O pacote *Atropos* aplica o teorema de Bayes, ou seja, depende de um modelo de verosimilhança e de uma probabilidade *a priori*, modelizada com MRF. Assim, definiu-se que o modelo a utilizar seria a distribuição gaussiana, para três classes, e que o MRF seria aplicado no

plano axial, com raio de 1 *voxel*, e fator de suavização de 0,5, aplicando-se somente na sequência T1, dado que ao utilizar as três a segmentação falhava, talvez porque o número de gaussianas não era adequado.

Atlas – um atlas de probabilidades de tecidos é uma imagem que em cada *voxel* tem a probabilidade desse mesmo *voxel* pertencer a um tecido. Tem sido utilizado em abordagens estatísticas, uma vez que permite incorporar informação espacial durante a segmentação. Neste mesmo sentido, utilizou-se um atlas, por cada classe, como característica.

Na determinação do atlas é necessário aplicar um processo de alinhamento de imagens, através da minimização de uma função de custo, denominado registo. Começou-se por seleccionar um sujeito de treino, aleatoriamente, para o qual se registou as sequências T1 dos restantes. Utilizando-se a matriz de transformação resultante, registou-se o *ground truth* de cada sujeito para o mesmo espaço comum, tendo sido utilizado para calcular a probabilidade de cada tecido em cada *voxel*, calculando a média de cada classe, resultando em três atlas, um por cada tecido. Assim, para a utilização posterior dos atlas, registou-se a imagem T1 de referência para o sujeito de interesse, e utilizou-se a matriz de transformação para transformar os atlas para esse espaço. Com estas características consegue-se introduzir informação espacial. Nas Figura 7.9(g-i) encontram-se os atlas de probabilidades para LCR, SC e SB, respetivamente.

Foi utilizado o registo afim, pois este permite que exista alguma deformação da imagem, o que faz com que exista alguma adaptabilidade às variabilidades anatómicas existentes entre sujeitos; isto deve-se ao facto de este tipo de registo utilizar 12 graus de liberdade (3 para translação, 3 para rotação, 3 para escala e 3 para deformação). Já a função de custo seleccionada foi a *mutual information*, pois estabelece uma relação estatística entre as intensidades das imagens, sendo mais flexível do que outras abordagens que pressupõe que sejam próximas. O problema de registo está sempre associado a interpolação, uma vez que é necessário estimar as intensidades dos *voxels* ao fazer-se as transformações espaciais. Assim sendo, utilizou-se interpolação trilinear quando se registou as imagens T1, por ser uma abordagem em que os valores interpolados são contínuos, enquanto para o registo do *ground truth* e dos atlas utilizou-se o vizinho mais próximo, por ser um método discreto e por manter os valores das imagens constantes, isto é, no caso dos *ground truth* não acrescenta casas decimais, nem altera os valores das probabilidades [91]–[93]. No pacote de *software* FSL existe a ferramenta *FMRIB's Linear Image Registration Tool* (FLIRT), tendo sido utilizada para o processo de registo [91], [92].

Módulo do gradiente – esta característica foi incluída com o propósito de adicionar informação nas fronteiras entre tecidos, tendo sido calculada com o *Insight Toolkit*. Para que a fronteira entre o LCR e o fundo da imagem não sobressaísse, criou-se uma máscara do encéfalo que foi dilatada, iterativamente, com morfologia matemática. Nas zonas dilatadas colocou-se a média dos *voxels* mais próximos, e no final retirou-se a zona acrescentada à imagem do gradiente. Calculou-se o gradiente para as três sequências.

Apesar de se utilizar uma abordagem com várias sequências, quando se segmentou as imagens MPRAGE dos EVR apenas essa foi utilizada, o que impediu o cálculo das diferenças entre sequências, reduzindo o número de características para 20. Não se utilizou a FLAIR disponível, uma vez que a mesma foi adquirida com distância entre *slices* de 6 mm, considerando-se elevado para se utilizar uma abordagem de interpolação.

7.2.2.2. Classificação e pós-processamento

Depois de extrair as características, classificou-se os sujeitos de teste com RF. Foi utilizada a implementação WiseRF³, pois os *benchmarks* disponibilizados mostram melhores características do que outras alternativas existentes (Tabela 7.1).

Ao utilizar-se a referida implementação teve que se extrair todas as características antes da classificação. Além disso, o resultado é calculado pela média das DTs, uma vez que apenas esta abordagem estava disponível. Utilizou-se *leave-one-out* para avaliar o efeito das características, e o *grid search* para procurar os melhores hiperparâmetros para a RF.

O pós-processamento consistiu em dois procedimentos. Verificou-se que alguns *voxels* de LCR tinham intensidades negativas, sendo excluídos da máscara do encéfalo e considerados *voxels* de fundo. Portanto, utilizou-se o conjunto complementar do fundo, e considerou-se que o fundo deveria formar um componente ligado com muitos *voxels*, e aqueles componentes que fossem constituídos por menos de 15 elementos foram incluídos na classe de LCR. Por outro lado, considerou-se que era pouco coerente se existissem regiões de uma classe com menos de um pequeno número de *voxels* (validado no capítulo de resultados) no meio de outros tecidos, assim, detetou-se esses componentes e substituiu-se a sua classe pela moda da vizinhança.

No final, a extração da SB cerebral é trivial, bastando utilizar a *label* 3 da classificação como máscara.

³ Disponibilizada na sua versão académica pela empresa *Wise.io*

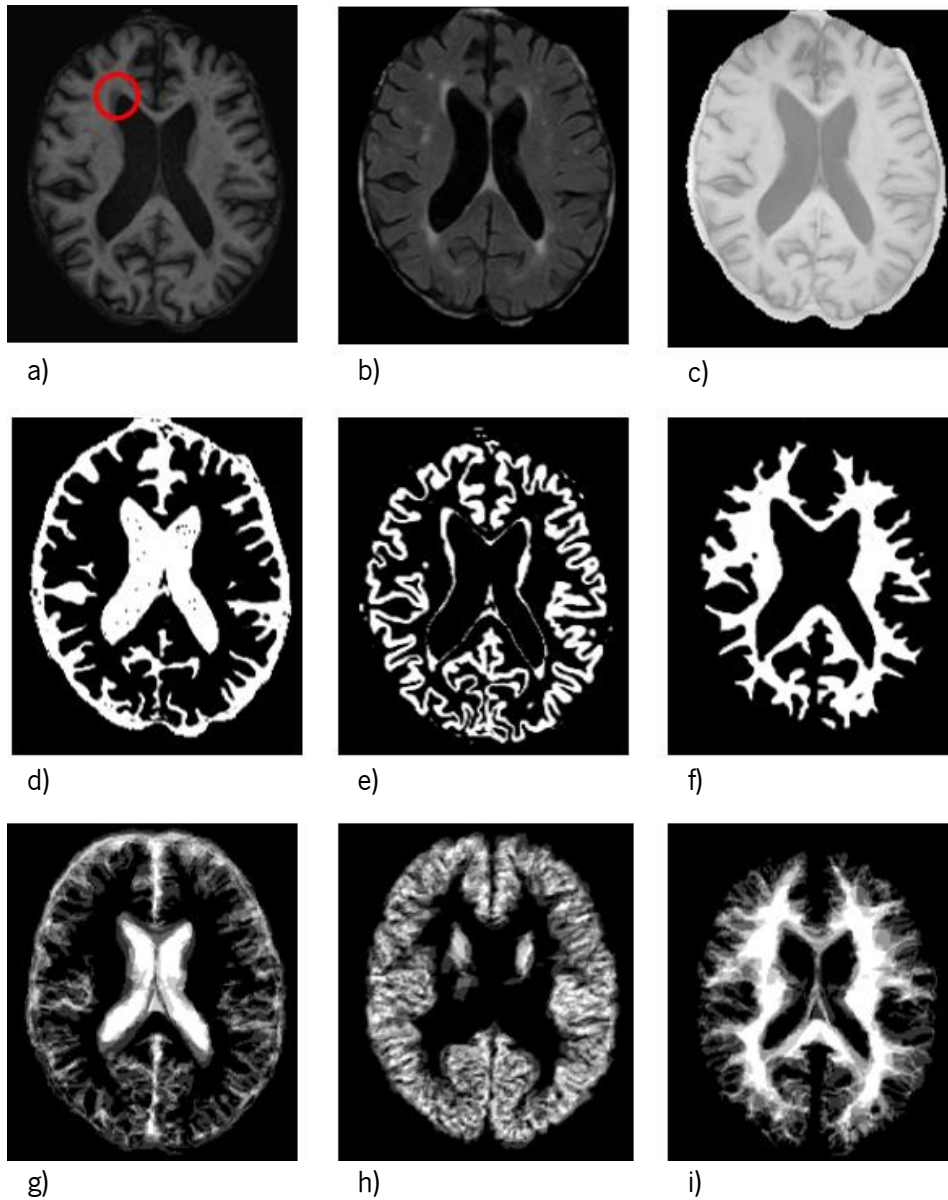


Figura 7.9 - Algumas características. a-c) seqüências T1w, FLAIR e T1w IR, respectivamente. d-f) probabilidades *a posteriori* para LCR, SC e SB, respectivamente. g-i) atlas de probabilidades para LCR, SC e SB. Em a) observa-se um círculo vermelho a marcar uma região de lesão da SB, que foi incluída na SC da imagem e) pelo método *Atropos*.

Tabela 7.1 - Comparação entre diferentes implementações de *Random Forests*. Consultado em [94].

	WiseRF	Python sklearn	H2O	R
Erro no teste	2,6%	2,6%	2,7%	3%
Tempo de treino (s)	29	105	71	960
Tempo de teste (s)	0,9	7,1	13,4	2,5

7.3. Detecção dos Espaços de Virchow-Robin

O algoritmo de deteção de EVR dilatados implementado foi o desenvolvido por *Descombes et al.*, em 2004. Este pressupõe que as estruturas são cilíndricas, com diâmetro entre 1 mm e 3 mm, e comprimento entre 3 mm e 15 mm, e que se pode modelizar a sua distribuição através de um MPP. Finalmente, otimizando o modelo com o algoritmo RJMCMC embebido em *simulated annealing* pretende-se calcular a configuração mais provável das estruturas [24]. Na Figura 7.10 pode-se rever a *pipeline* utilizada para a deteção automática de EVR.

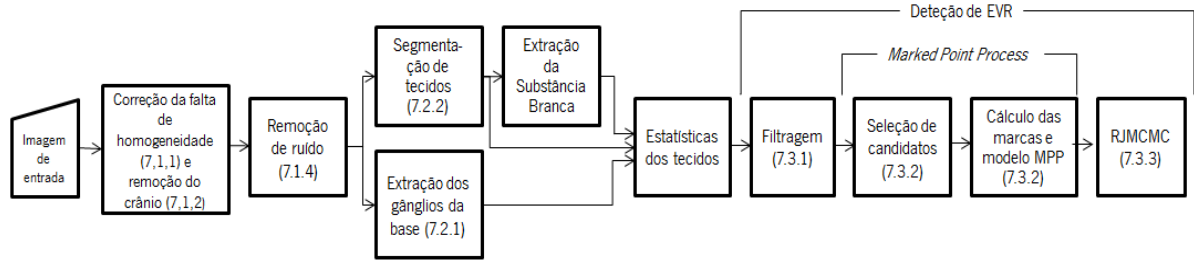


Figura 7.10 - *Pipeline* para a deteção automática de Espaços de Virchow-Robin.

7.3.1. Filtragem

Começou-se por aplicar três filtros para realçar os EVR, em cada plano do espaço (u, v, w) .

Os filtros são aplicados sobre o *voxel* central, c , os seus 8 vizinhos mais próximos $t_1 \in \mathcal{N}_u^1(c)$ e os 16 vizinhos que rodeiam os anteriores, $t_2 \in \mathcal{N}_u^2(c)$, estando definido em relação à direção do espaço u , sendo semelhante para as restantes. Pode-se observar os tipos de vizinhança na Figura 7.11(a).

As intensidades dos EVR e vizinhança apresentam três características: o *voxel* central tem baixa intensidade em T1, os *voxels* da vizinhança tem intensidade mais elevadas e contrastam com o central. Assim, a intensidade do *voxel* central (i_c) é reduzida em T1 por causa do LCR do espaço, sendo realçado pelo filtro $F^{escuro}(i_c)$ (Figura 7.11(b)). Os tecidos envolventes (SC e SB) têm intensidades mais elevadas, podendo até aumentar a do próprio espaço devido a efeitos de volume parcial, codificando-se esta propriedade no filtro $F^{claro}(\min(i_t, t \in \mathcal{N}_u^2(c)))$ (Figura 7.11(c)). Por fim, os mesmos tecidos contrastam com EVR, estando associado ao filtro $F^{contraste}(\sum_{t \in \mathcal{N}_u^2(c)} i_t / 16)$ (Figura 7.11(d)) [24].

Os filtros são construídos tendo em conta a média e o desvio padrão da SC, SB e LCR, portanto estas regiões devem estar segmentadas. Os autores utilizaram crescimento de regiões, mas considerou-se que esse tipo de abordagem deveria ser supervisionado por um utilizador, para garantir que não há extravasamento das regiões para áreas de diferentes tecidos, o que não se adequava ao objetivo de se desenvolver um método totalmente automático para a deteção de EVR dilatados. Assim sendo, substituiu-se essa abordagem pelo método desenvolvido para a extração de SB, apresentado na secção 7.2.2. , utilizando-se a segmentação do *Freesurfer* para remover o cerebello e tronco encefálico, de modo a reduzir os erros no cálculo das estatísticas dos tecidos.

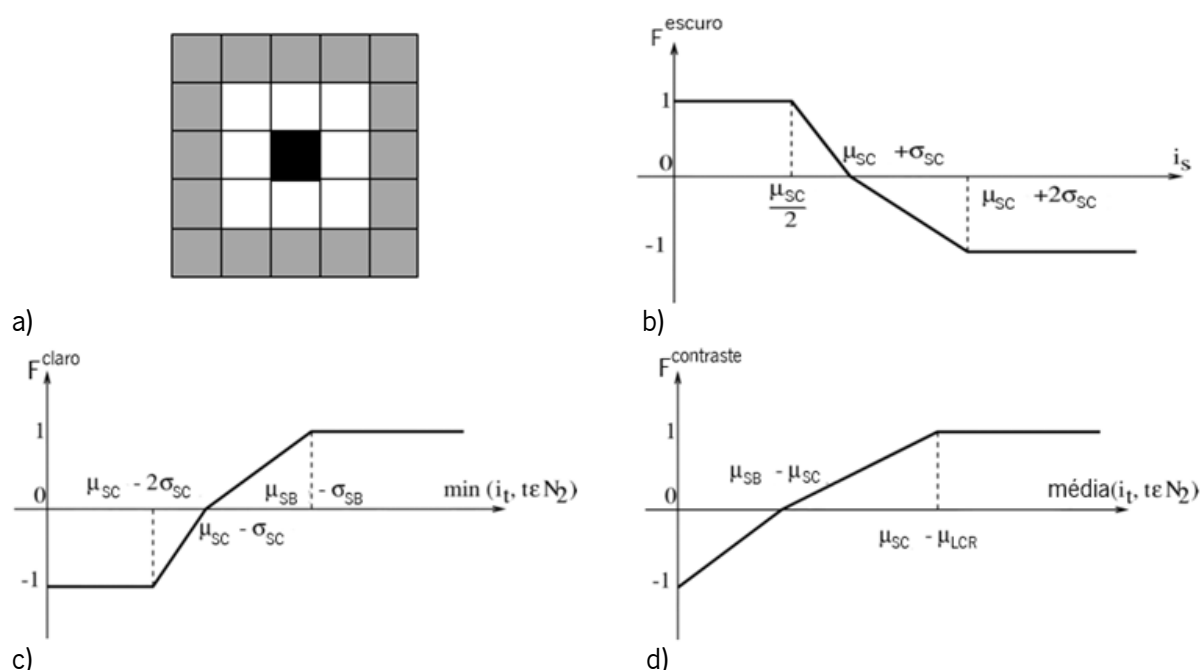


Figura 7.11 – Filtros de propriedades dos Espaços de Virchow-Robin. Em a) observa-se a área de aplicação dos filtros, nomeadamente o *voxel* negro é o central, a branco é a vizinhança \mathcal{N}_u^1 , enquanto a cinzento está marcada a vizinhança \mathcal{N}_u^2 . De b-d) encontram-se representados os filtros F^{escuro} , F^{claro} e $F^{contraste}$, respetivamente.

Adaptado de [24].

As zonas de resposta dos filtros correspondente aos segmentos de reta que não são constantes, isto é, que tem inclinação, tem o propósito de conferir alguma flexibilidade, pois são realçadas zonas que se afastem um pouco das propriedades referidas anteriormente, nomeadamente por causa de efeito de volume parcial. Por outro lado, também podem ser úteis para compensar erros de segmentação dos tecidos, que façam com que as estatísticas calculadas não sejam ótimas, ou exista grande variabilidade. Verifica-se, também, analisando a Figura 7.11(c), que o modelo parece estar especialmente projetado para detetar EVR na SB, no

entanto, valores da vizinhança correspondentes a SC também resultam em valores positivos dos filtros, pelo que se espera que o algoritmo também funcione nos gânglios da base.

Os três filtros são combinados num só como descrito na Equação (7.9) para a direção u (igual para as restantes direções), que reúne a resposta mais fraca deles. Além disso, previne que sejam realçados vários *voxels* para o mesmo local, ao tentar forçar que apenas se registre a resposta para o elemento menos intenso de uma vizinhança, a não ser que reúna as três propriedades assumidas para os *voxels* candidatos a EVR [24].

$$F_u(c) = \begin{cases} f_u(c), & \text{se } \forall t \in \mathcal{N}_u^1(c), i_c \leq i_t \\ \min(0, f_u(c)), & \text{caso contrário} \end{cases}$$

$$f_u(c) = \min \left(F^{\text{escuro}}(i_c), F^{\text{claro}} \left(\min(i_t, t \in \mathcal{N}_u^2(c)) \right), F^{\text{contraste}} \left(\frac{\sum_{t \in \mathcal{N}_u^2(c)} i_t}{16} \right) \right) \quad (7.9)$$

7.3.2. Marked Point Process

A distribuição dos EVR dilatados não é aleatória na região de interesse, existindo determinadas orientações, entre aqueles que estão mais próximos, que são mais prováveis do que outras. Além disso, a sua geometria é conhecida, visto ter-se assumido que se iria detetar aqueles que apresentam forma aproximadamente cilíndrica. Estas propriedades podem ser modelizadas através de um MPP, semelhante ao apresentado na secção 4.2.1. constituído por dois termos (Equação (7.10) [24].

$$h(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x}) \quad (7.10)$$

onde $f(\mathbf{x})$ é o termo *a priori*, sendo utilizado um modelo de Candy, que toma em consideração a geometria e a distribuição dos EVR. $g(\mathbf{x})$ é o termo proveniente dos dados.

O volume da região de interesse, V , é restrito aos pontos x_i , com $x_i \in V$, que obedecem à condição $\max (F_u(x_i), F_v(x_i), F_w(x_i)) \geq 0$. A cada ponto está associada uma marca, \vec{l}_i , que corresponde ao comprimento e orientação de cada candidato a EVR, sendo este um vetor em \mathcal{R}^3 , com a limitação $\|\vec{l}_i\| \in [\rho_{\min} = 3, \rho_{\max} = 15]$. Assim sendo, se $\vec{s}_i = (x_i, \vec{l}_i)$ representar cada um dos pontos marcados, a configuração de n pontos marcados resultante é $\mathbf{x} = \{\vec{s}_1 = (x_1, \vec{l}_1), \dots, \vec{s}_n = (x_n, \vec{l}_n)\}$ [24].

Para se definir cada uma das marcas, percorreu-se cada um dos pontos x_i . Em cada um deles, verificou-se qual a direção do espaço que resultou na maior resposta do filtro da Equação (7.9). Então utilizou-se o resultado da filtragem segundo essa mesma direção para procurar l_i . Assim, começou-se a procura no ponto imediatamente a seguir e os seus 8 vizinhos, segundo essa direção. Em seguida verificou-se qual deles era aquele que resultou num valor mais elevado da filtragem, e moveu-se a procura para esse ponto, repetindo-se o procedimento anterior até se atingir um *slice* onde os nove elementos de procura tivessem resposta negativa. Note-se que em cada iteração se move um *slice* segundo a direção para a qual a resposta do primeiro elemento foi máxima. Desta forma, consegue-se saber o comprimento do candidato a EVR, e a sua direção, segundo cada eixo.

O termo *a priori* é definido com densidade relacionada com uma medida de Poisson, de forma semelhante ao modelo de Candy, e modeliza o comprimento dos candidatos, e as possíveis interações entre eles. Na Equação (7.11) encontra-se o modelo utilizado.

$$f(x) \propto \beta^n \prod_{i \in \{1, n\}} q_1(\|\vec{l}_i\|) \prod_{i, j \in \{1, n\}: i \sim j} q_2(\vec{s}_i, \vec{s}_j) \quad (7.11)$$

onde β é o parâmetro de densidade de Poisson, q_1 é o termo que representa o *prior* do comprimento dos candidatos, q_2 é o termo de interação entre candidatos vizinhos com relação de vizinhança $i \sim j$.

O termo q_1 é definido pela Equação (7.12) [24].

$$q_1(\|\vec{l}_i\|) = \exp\left(-A \left(\frac{\rho_{max} - \|\vec{l}_i\|}{\rho_{max} - \rho_{min}}\right)^2\right) \quad (7.12)$$

Considerou-se três possíveis interações entre pontos marcados vizinhos, duas de penalização, e uma de favorecimento. Para as definir foi necessário considerar a silhueta dos segmentos, $S(\vec{s})$, tratando-se da projeção dos mesmos no vetor 3D que forma a imagem de RM [24].

Deste modo, penalizou-se interseções entre segmentos (interação *hard core*), como o exemplo da Figura 7.12(a), utilizando a Equação (7.13) [24].

$$q_2^{hc}(\vec{s}_i, \vec{s}_j) = \exp(-B), \text{ se } S(\vec{s}_i) \cap S(\vec{s}_j) \neq \emptyset \quad (7.13)$$

Favoreceu-se casos em que os candidatos estão aglomerados (*clustering*), e seguem direções semelhantes, como na Figura 7.12(b). Considerando o comprimento segundo as três direções do espaço, l_u , l_v e l_w , a condição de proximidade e paralelismo é definida pela

Equação (7.14), para a direção u (recorde-se que se utiliza esta direção como exemplo, mas deve-se verificar para todas). Caso se verifique a condição, calcula-se o termo de interação com a Equação (7.15) [24].

$$\begin{cases} |l_{u_i}| > \max(|l_{v_i}|, |l_{w_i}|) \\ |l_{u_j}| > \max(|l_{v_j}|, |l_{w_j}|) \\ l_{u_i} \cdot l_{u_j} > 0 \end{cases} \quad (7.14)$$

$$q_2^{cl}(\vec{s}_i, \vec{s}_j) = \exp(+C \min |l_{u_i}, l_{u_j}|), \text{ se } |x_{u_i} - x_{u_j}| < c \quad (7.15)$$

Por fim, o terceiro tipo penaliza configurações com segmentos próximos, mas direções diferentes (Equação (7.16)), tratando-se de interações repulsivas, tal como esquematizado na Figura 7.12(c). Estas relações são mutuamente exclusivas relativamente às de *clustering* [24].

$$q_2^{rep} = \exp(-D), \text{ se } \|x_i - x_j\| < d \quad (7.16)$$

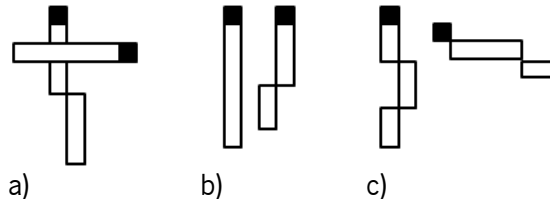


Figura 7.12 - Tipos de interação entre Espaços de Virchow-Robin. O ponto preenchido a preto representa o *voxel* inicial de cada segmento. Em a) é uma interação *hard core*, em b) de *clustering* e em c) de repulsão.

A Equação (7.17) define o termo proveniente dos dados. Este procura dar maior peso às configurações cujos candidatos se adequem mais às propriedades das intensidades dos EVR, e da sua vizinhança [24].

$$g(x) \propto \prod_{i \in \{1, n\}} r(\vec{s}_i) \quad (7.17)$$

O termo $r(\vec{s})$ é constituído por duas partes, na primeira toma em consideração a resposta aos filtros já descritos, sendo proporcional a esta (Equação (7.9)); na segunda avalia os *voxels* e a vizinhança \mathcal{N}^1 que se encontram imediatamente a seguir às extremidades de cada candidato, considerando que estes devem ter intensidades mais elevadas do que os EVR, pois já não são constituídos por LCR, através da função $F_u^{extremos \text{ claros}}(\vec{s})$, escrita relativamente à direção u , representada na Figura 7.13 [24]. Observa-se que os EVR da SB são beneficiados com valores mais elevados, mas mais uma vez, nos gânglios da base, ao se considerar um local de SC, todos saem afetados de igual forma, esperando-se que o método seja eficaz.

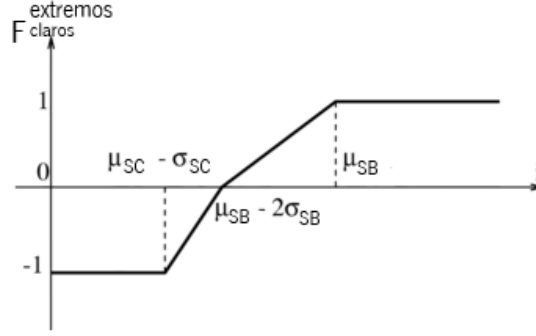


Figura 7.13 - Função para a avaliação da intensidade dos *voxels* e da vizinhança \mathcal{N}^1 imediatamente a seguir aos extremos dos candidatos a Espaços de Virchow-Robin. Adaptado de [24].

Combinou-se os resultados da aplicação da função $F_u^{extremos\ claros}$ a ambos os extremos dos pontos marcados através da Equação (7.18), descrita em relação à direção u , mas que é semelhantes para as restantes direções. Esta também apresenta mais alguns detalhes de utilização da função, considerando ainda a possibilidade de o candidato ter dois sentidos segundo um dado eixo, positivo ou negativo [24].

$$\begin{aligned}
 a_1 &= F^{extremos\ claros} \left(\min \left(i_{x+\vec{l}+(1,0,0)}, i_t, t \in \mathcal{N}_u^1(i_{x+\vec{l}+(1,0,0)}) \right) \right) \\
 a_2 &= F^{extremos\ claros} \left(\min \left(i_{x+(-1,0,0)}, i_t, t \in \mathcal{N}_u^1(i_{x+(-1,0,0)}) \right) \right) \\
 b_1 &= F^{extremos\ claros} \left(\min \left(i_{x+(1,0,0)}, i_t, t \in \mathcal{N}_u^1(i_{x+(1,0,0)}) \right) \right) \\
 b_2 &= F^{extremos\ claros} \left(\min \left(i_{x+\vec{l}+(-1,0,0)}, i_t, t \in \mathcal{N}_u^1(i_{x+\vec{l}+(-1,0,0)}) \right) \right)
 \end{aligned} \tag{7.18}$$

$$F_u^{extremos}(\vec{s}) = \begin{cases} a_1 + a_2, & \text{se } \vec{l}_u > 0 \\ b_1 + b_2, & \text{se } \vec{l}_u \leq 0 \end{cases}$$

Portanto, já se apresentou todos os elementos para calcular $r(\vec{s})$. Para tal, deve-se descobrir, para cada candidato, a sua direção no espaço para que a vizinhança \mathcal{N}^1 utilizada na função $F^{extremos\ claros}$ seja a correta. Na Equação (7.19), encontra-se a expressão de cálculo para a direção u , sendo semelhante para as restantes [24].

$$r(\vec{s}) = \exp \left(E \left[\sum_{t \in S(\vec{s})} F_u(t) + \frac{n_{\vec{s}}}{2} F_u^{extremos}(\vec{s}) \right] \right) \tag{7.19}$$

se $|l_u| \geq \max(|l_v|, |l_w|)$

onde $n_{\vec{s}}$ é o número de *voxels* em $S(\vec{s})$.

Na Tabela 7.2 encontram-se os parâmetros utilizados em [24] no modelo MPP para EVR.

Tabela 7.2 - Parâmetros do modelo *Marked Point Process* para Espaços de Virchow-Robin.

Parâmetro	Valor	Parâmetro	Valor
β	1	D	60
A	10	E	80
B	10^{10}	c	5
C	20	d	3

7.3.3. Otimização com RJMCMC

Definido o modelo MPP, $h(\mathbf{x})$, tornou-se necessário otimizá-lo para que fosse encontrada a configuração de EVR mais provável. Uma vez que o número de EVR é desconhecido, e varia, utilizou-se o algoritmo RJMCMC para simular as configurações. Para se descobrir o máximo global do modelo, aplicou-se *simulated annealing*, passando a amostrar-se a distribuição $h_T(\mathbf{x}) = h(\mathbf{x})^{\frac{1}{T}}$, tendo os autores utilizado um esquema de arrefecimento geométrico, na forma $T_{i+1} = \theta T_i$ [24]. No entanto, apenas foi referido que $0 < \theta < 1$ pelo que na ausência de marcações manuais não se pôde procurar o melhor θ . Assim sendo, optou-se por alterar o esquema de arrefecimento pelo arrefecimento exponencial da Equação (5.21), com $T_0 = 1$ e $C = 0,8$, pois foi com este valor que se obteve a estimativa do máximo mais próximo do verdadeiro na experiência da Figura 5.8, ainda que se desconheça o seu real comportamento se existirem mais modas, se estas estiverem mais próximas e em dimensões superiores.

Para o RJMCMC considerou-se cinco tipos de movimentos; já que em [24] não era referido qual a probabilidade de cada um, definiu-se que seria igual para todos, com o valor 0,2. A cada movimento está associado uma razão de aceitação, α , com basa na qual se decide se se aceita a nova configuração, \mathbf{x}' , ou se mantém a antiga, \mathbf{x} .

O primeiro tipo de movimento, e o essencial, consistiu em dois subtipos, o *birth* e o *death*, tendo-se considerado igual probabilidade para ambos, correspondendo à probabilidade global de 0,1. No *birth* propõe-se, uniformemente, um segmento, a partir do conjunto de todos os candidatos a EVR, enquanto no *death* se escolhe, uniformemente, um candidato do conjunto de EVR já aceites para ser removido [24]. Teve-se o cuidado de evitar propor um candidato quando já se aceitou todos os que estariam disponíveis, e não aceitar o movimento *death* se o

conjunto de segmentos aceites estivesse vazio. Nas Equações (7.20) e (7.21) encontram-se as razões de aceitação para *birth* e *death*, respetivamente.

$$\alpha_{birth} = \frac{P_{death}}{P_{birth}} \times \frac{h_T(\mathbf{x}')}{h_T(\mathbf{x})} \times \frac{4\pi^2(\rho_{max} - \rho_{min})V}{n + 1} \quad (7.20)$$

$$\alpha_{death} = \frac{P_{birth}}{P_{death}} \times \frac{h_T(\mathbf{x}')}{h_T(\mathbf{x})} \times \frac{n}{4\pi^2(\rho_{max} - \rho_{min})V} \quad (7.21)$$

O movimento *segment move* corresponde ao segundo tipo. Neste realiza-se um *death* e um *birth* em simultâneo, isto é, escolhe-se, uniformemente, um segmento do conjunto de candidatos aceites para remover, e sugere-se um novo candidato. Este movimento ajuda a atingir-se a convergência mais rapidamente [24]. A sua razão de aceitação está escrita na Equação (7.22)

$$\alpha_{segm.move} = \frac{h_T(\mathbf{x}')}{h_T(\mathbf{x})} \quad (7.22)$$

O terceiro movimento, *end point move*, consiste na seleção uniforme de um segmento da configuração, e na mudança do seu último *voxel* para um local da sua vizinhança, escolhido uniformemente, para que o segmento se enquadre melhor nas propriedades dos EVR [24]. A sua razão de aceitação é a da Equação (7.23).

$$\alpha_{end pt.move} = \frac{h_T(\mathbf{x}')}{h_T(\mathbf{x})} \quad (7.23)$$

O quarto tipo de movimento também é constituído por dois subtipos, o *split* e o *merge*, cada um com probabilidade global de 0,1. O *merge* foi utilizado para juntar dois segmentos da configuração, com o objetivo de não serem detetados dois EVR, quando, na realidade, seriam o mesmo, mas parcialmente detetado. Para que isso aconteça, o último *voxel* de um dos segmentos deve estar próximo do primeiro do outro, satisfazendo uma das condições da Equação (7.24). Assim, se existirem dois segmentos $\vec{s}_1 = (x_1, \vec{l}_1)$ e $\vec{s}_2 = (x_2, \vec{l}_2)$, a sua junção é $\vec{s}^{nov\vec{o}} = (x_1, \vec{l}^{nov\vec{o}} = (x_1, x_2 + \vec{l}_2))$ se for verificada a primeira condição da Equação (7.24), ou $\vec{s}^{nov\vec{o}} = (x_1, \vec{l}^{nov\vec{o}} = (x_2, x_1 + \vec{l}_1))$, no caso de a segunda ser verdadeira [24].

$$\begin{aligned} \max(|x_{u_2} - x_{u_1} - l_{u_1}|, |x_{v_2} - x_{v_1} - l_{v_1}|, |x_{w_2} - x_{w_1} - l_{w_1}|) &\leq D_m \\ \max(|x_{u_2} + l_{u_2} - x_{u_1}|, |x_{v_2} + l_{v_2} - x_{v_1}|, |x_{w_2} + l_{w_2} - x_{w_1}|) &\leq D_m \end{aligned} \quad (7.24)$$

com $D_m = 1$.

O movimento *split* separa um segmento da configuração em dois. Considere-se o segmento da configuração $\vec{s} = (x, \vec{l}) = \overline{xy}$, com $y = x + \vec{l}$. Os pontos candidatos para a separação, p , são aqueles que obedecem a $\|\overline{xp}\| \geq \rho_{min}$ e $\|\overline{py}\| \geq \rho_{min}$, procedendo-se à escolha de um deles, uniformemente, e dividindo-se o segmento (Figura 7.14(a)). O ponto final da primeira parte pode ser escolhido entre os 8 vizinhos mais próximos de p , no plano (Figura 7.14(b)), enquanto o elemento inicial do segundo novo segmento é escolhido entre o volume que rodeia p , tendo em consideração que o comprimento da aresta desse cubo está relacionado com o parâmetro D_m (Figura 7.14(c)) [24].

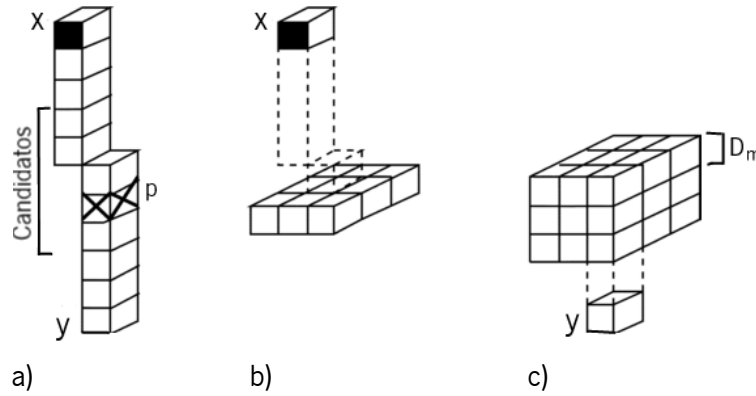


Figura 7.14 - Movimento *split*. Em a) encontra-se o segmento selecionado, respetivos pontos candidatos para a divisão e o ponto escolhido, enquanto que em b) e c) se demonstra os *voxels* candidatos para o ponto final do novo segmento correspondendo à primeira parte do original, e os candidatos para o ponto inicial da segunda parte, respetivamente. Adaptado de [24].

Nas Equações (7.25) e (7.26) pode-se consultar as razões de aceitação para o *split* e *merge*, respetivamente.

$$\alpha_{split} = \frac{P_{merge}}{P_{split}} \times \frac{h_T(\mathbf{x}')}{h_T(\mathbf{x})} \times \frac{(2D_m + 1)^5 (|l_u| - 2\rho_{min})}{n_{aceites} + 1} \quad (7.25)$$

$$\alpha_{merge} = \frac{P_{split}}{P_{merge}} \times \frac{h_T(\mathbf{x}')}{h_T(\mathbf{x})} \times \frac{n_{aceites} + 1}{(2D_m + 1)^5 (|l_u| - 2\rho_{min})} \quad (7.26)$$

Por fim, o movimento *return* inverte a direção de um segmento do conjunto de candidatos aceites, escolhido uniformemente. O movimento *merge* apenas deixa fundir dois segmentos que tenham a mesma direção; assim, com o *return* pode-se fazer com que mais segmentos possam ser unidos [24]. A sua razão de aceitação corresponde à Equação (7.27).

$$\alpha_{return} = \frac{h_T(\mathbf{x}')}{h_T(\mathbf{x})} \quad (7.27)$$

Os autores testaram o algoritmo em 37 sujeitos, divididos em três classes: os que apresentavam reduzido número de EVR, moderado e elevado. Testando em todo o conjunto, reportou-se correlação de 0,77, utilizando regressão linear, entre as contagens do algoritmo e as dos especialistas. Comparando com as marcações manuais na classe de sujeitos com maior número de EVR, o coeficiente de correlação interclasse foi de 0,87, a taxa de falsos positivos foi de 13% e de falsos negativos de 2,7% [24].

7.4. Aplicação para visualização de Espaços de Virchow-Robin

Após a deteção automática dos EVR, pode ser útil visualizá-los de forma a verificar se o processo foi bem sucedido, analisar a sua distribuição na região de interesse, a sua estrutura e verificar se a região de interesse está corretamente segmentada.

Nesse sentido, desenvolveu-se uma aplicação em *MeVisLab*, cujo painel principal se pode observar na Figura 7.15. Na secção B seleciona-se a imagem da região de interesse, a imagem com os EVR detetados com intensidade binária e a imagem onde cada um tem uma intensidade (ou *label*) diferente. É verificado o número de espaços detetados, fornecendo-se essa informação. Na secção D, de cima para baixo, encontram-se os planos axial, coronal e sagital da imagem de RM, com os EVR sobrepostos a vermelho. Caso se deseje analisar melhor as intensidades dos mesmos, pode-se retirar as marcações vermelhas, clicando na opção *on*, da secção B. Ao clicar com o rato num *voxel* pertencente ao EVR, a estrutura ao qual pertence é detetada sabendo a *label* do ponto selecionado, e é reconstruída tridimensionalmente na secção C, onde se pode ver a sua forma, por exemplo para verificar se é muito tortuoso, podendo ser útil para identificar se é, de facto, um espaço dilatado; o elemento correspondente ao *voxel* selecionado é marcado (no caso da imagem com uma cruz verde).

No painel *3D* (Figura 7.16) pode-se verificar a região de interesse utilizada, através da sua reconstrução, e a distribuição dos EVR no espaço, permitindo obter-se uma ideia melhor da forma como estão distribuídos. Os quadrados a roxo podem ser movidos, e ao interseccionar os objetos fica desenhado na sua superfície o corte 2D dessa região.

Por fim, no painel *Overlay_3D* (Figura 7.17) observa-se os EVR sobrepostos à reconstrução 3D da região de interesse. Assim, utilizando os quadrados de corte, pode-se perceber de que forma as estruturas percorrem o encéfalo.



Figura 7.15 - Janela principal da aplicação para visualização dos Espaços de Virchow-Robin na região de interesse. A - seleção do painel, B - secção de seleção da imagem e opções, C - visualizador 3D da estrutura do EVR, D - visualizador da imagem de RM.

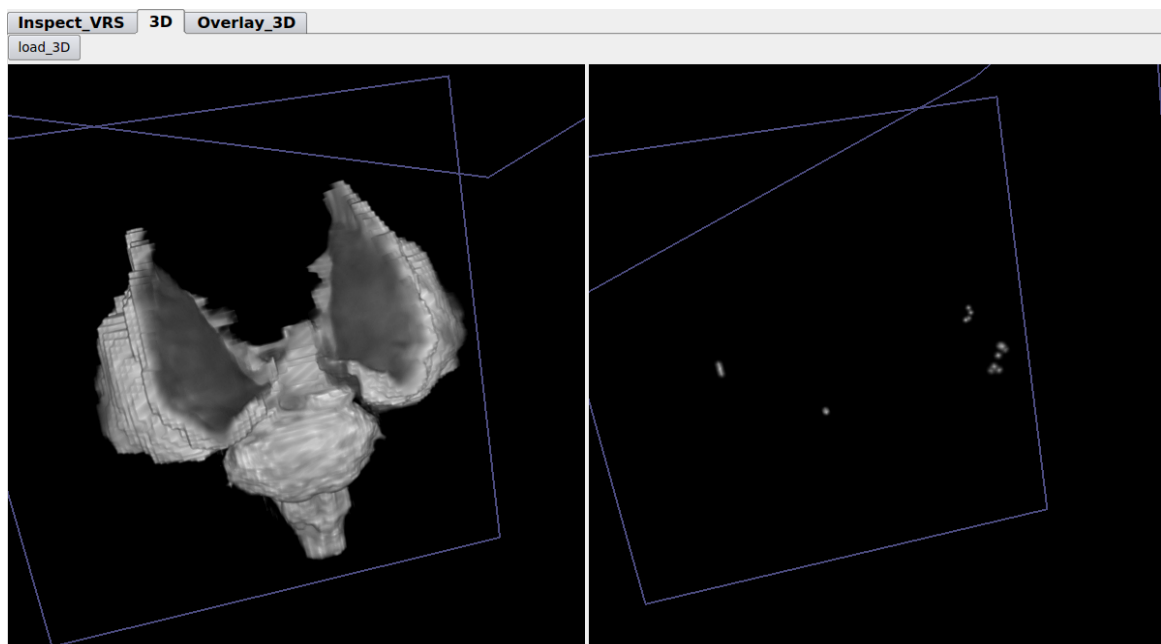


Figura 7.16 - Painel 3D da aplicação. À esquerda a reconstrução da região de interesse, e à direita a configuração dos EVR detetados, no espaço.

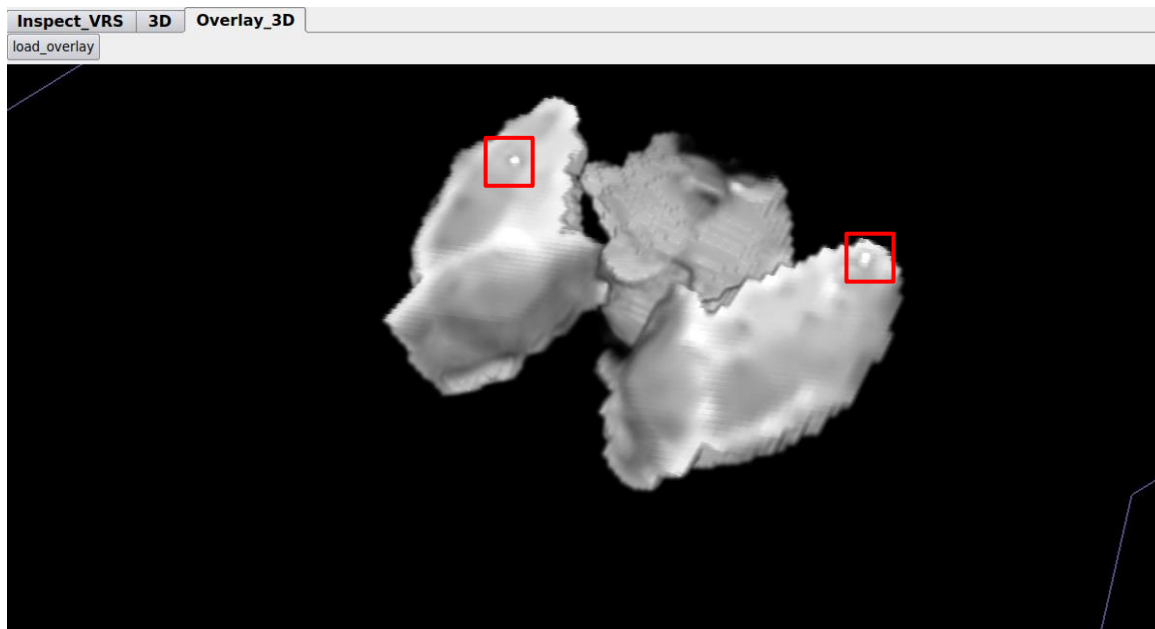


Figura 7.17 - Painel *Overlay_3D*. Na região a vermelho marca-se um EVR sobreposto à região de interesse.

7.5. Sumário

O processamento de imagens de RM requer a utilização de vários pré-processamentos, de forma a facilitar o processo, ou mesmo torná-los possíveis. No sistema implementado, considerou-se que os mais importantes seriam a correção da falta de homogeneidade, para que o mesmo tecido numa imagem tivesse uma gama de intensidade uniforme, a remoção do crânio, para reduzir o volume de processamento, a normalização das intensidades entre diferentes sujeitos, que toma especial importância em algoritmos supervisionados, e a remoção do ruído.

A extração da região de interesse é importante para dar mais significados aos EVR detetados, mas também diminui a carga computacional. Para a extração dos gânglios da base utilizou-se o *Freesurfer*, enquanto para a extração de SB foi desenvolvido um método original de segmentação de tecidos cerebrais, utilizando RFs.

Por fim, implementou-se o algoritmo de deteção de EVR [24], cujo ponto forte é a modelização dos EVR e as suas interações através de um MPP. Para se descobrir a configuração mais provável de EVR, otimizou-se o modelo, para cada sujeito, com o algoritmo RJMCMC com *simulated annealing*.

Capítulo 8 Resultados e discussão

Neste capítulo irá começar-se por apresentar as características das bases de dados MRBrainS, com a qual se validou o algoritmo de segmentação de tecidos do encéfalo, e a obtida na Escola de Ciências da Saúde, para a deteção de EVR.

Depois de se referir as métricas utilizadas na avaliação, irá expor-se e discutir os resultados obtidos na segmentação de LCR, SC e SB, e na deteção dos EVR.

8.1. Base de dados MRBrainS

A base de dados MRBrainS foi criada no contexto do *MICCAI Grand Challenge on MR Brain Image Segmentation 2013*, sendo constituída por cinco sujeitos de treino e quinze de teste, cuja idade é superior a 50 anos, sofrem de diabetes e têm elevado risco cardiovascular; além disso, apresentam diversos graus de atrofia e lesões da SB. Estas características aproximam-se dos dados disponíveis para a deteção de EVR.

Para cada paciente existia uma sequência T1w 3D (1mm x 1mm x 1mm), uma T1w (0,958mm x 0,958mm x 3mm), uma T1 IR (0,958mm x 0,958mm x 3mm) e uma FLAIR (0,958mm x 0,958mm x 3mm), adquiridas num *scanner* de 3 T. Para cada sujeito foram providenciadas duas marcações manuais para as sequências com distância entre *slices* de 3 mm; uma delas marcava o LCR, a SC e a SB, enquanto a outra continha o *ground truth* para SC cortical, gânglios da base, SB, lesões da SB, LCR no espaço extracerebral, ventrículos, cerebelo e tronco cerebral.

8.2. Base de dados para deteção de Espaços de Virchow-Robin

A base de dados utilizada para a deteção de EVR foi adquirida no Hospital de Braga, sendo disponibilizada para utilização na presente dissertação pela Escola de Ciências da Saúde.

Foram utilizados nove pacientes vítimas de AVCs, para os quais existiam, entre outras, as sequências MPAGE (1 mm x 1mm x 1 mm), FLAIR (0,49 mm x 0,49 mm x 6 mm) e T2 (1,2

mm x 1,2 mm x 1,2 mm). Na detecção dos EVR foi utilizada a MPRAGE, dado ser aquela cuja resolução era isotrópica de 1 mm.

8.3. Métricas de avaliação

Para a avaliação do sistema de segmentação de SC, SB e LCR, a base de dados MRBrainS utiliza três métricas: o coeficiente Dice, a distância de Hausdorff e a diferença absoluta do volume de cada classe.

A primeira métrica, o coeficiente Dice, mede a sobreposição entre a mesma classe segmentada pelo algoritmo implementado e pela marcação manual. Considerando que S representa a segmentação obtida, R a marcação manual, VP (verdadeiros positivos) o número de *voxels* corretamente classificados, FP (falsos positivos) a quantidade de *voxels* detetados como pertencentes à classe, mas que na verdade não são, e FN (falsos negativos) o número daqueles que não foram incluídos na classe pela segmentação, mas que deveriam ter sido, então pode-se calcular o coeficiente Dice utilizando a Equação (8.1) [53].

$$Dice = \frac{2|S \cap R|}{|S| + |R|} = \frac{2VP}{FN + 2VP + FP} \quad (8.1)$$

Esta métrica permite concluir acerca dos VP, FN e FP. Além disso, é uma das métricas mais populares na segmentação de RM do encéfalo, em conjunto com o Jaccard, podendo ser convertidas uma na outra. Entre os trabalhos que utilizaram Dice cita-se [9], [46], [49], [50], [53].

A segunda métrica, a distância de Hausdorff, mede o maior comprimento entre cada ponto de um conjunto e o ponto mais próximo de outro conjunto. Isto é, para cada *voxel* de uma classe da imagem segmentada, mede-se a distância entre este e o mais próximo da mesma classe no *ground truth*, sendo a distância de Hausdorff a maior das medidas [95]. No MRBrainS utiliza-se a distância de Hausdorff modificada, definida como a distância do 95° percentil.

A terceira métrica do MRBrainS mede a diferença absoluta do volume de cada classe.

8.4. Avaliação da segmentação de tecidos do encéfalo

Usando o *dataset* do MRBrainS, testou-se o impacto das características e dos parâmetros sobre a segmentação utilizando *leave-one-out*. A escolha desta abordagem deveu-se ao facto de o conjunto de treino apenas conter cinco sujeitos, pelo que a divisão em *K-folds*, com $K < 5$, faria com que existissem menos dados para a aprendizagem.

Uma vez que o objetivo era segmentar LCR, SC e SB, optou-se por treinar com essas classes, utilizando as três sequências com distância entre *slices* de 3 mm. Como cada treino foi realizado com quatro imagens, testando-se na quinta, não existia hipótese de uma das classes não estar presente em algum dos conjuntos, pelo que a utilização de estratificação não foi considerada necessária.

Os resultados foram calculados tal como no *challenge*, isto é, removendo-se o cerebelo e o tronco encefálico com uma máscara. Para as avaliações utilizou-se a métrica Dice.

8.4.1. Avaliação do método *Atropos*

Avaliou-se o método *Atropos* para se poder estabelecer uma comparação com o método proposto, mas também para otimizar o parâmetro de suavização do MRF, para que depois seja aplicado na extração da característica probabilidades *a posteriori*. Apenas foi utilizada a sequência T1, uma vez que determinar o número de gaussianas de uma abordagem multi-sequência não é trivial.

Observando a Figura 8.1, verifica-se que a segmentação do LCR piora à medida que se aumenta a suavização provocada pelo MRF, talvez porque se vão suavizando detalhes mais finos, como os sulcos mais estreitos que estariam preenchidos por líquido. Já a SB e a SC tem o seu pico de Dice para 0,5, o que indicia um compromisso entre as classes nos giros, onde a SB se torna mais estreita. Verifica-se, ainda, que o algoritmo é bastante robusto para o LCR, pois o seu desvio padrão para essa classe é sempre reduzido. Apesar de a diferença máxima entre os resultados ser de cerca de 0,01, existem diferenças visuais notáveis, tal como se pode perceber analisando a Figura 8.2. O aumento do parâmetro influencia a quantidade de pontos classificados isoladamente, tornando a segmentação mais consistente.

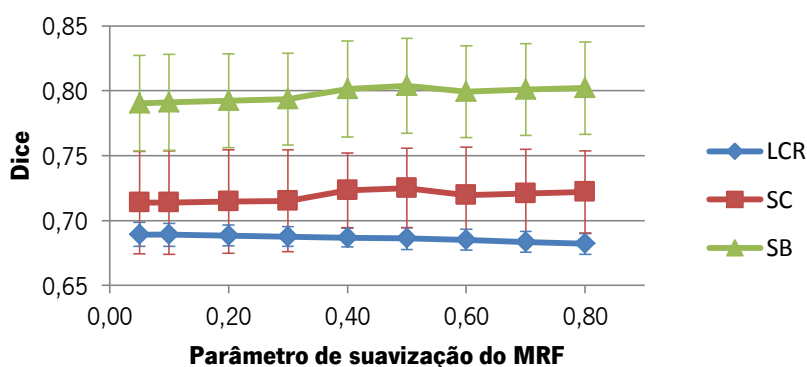


Figura 8.1 - Avaliação do algoritmo *Atropos*. Fez-se variar o parâmetro de suavização do MRF e mediu-se o seu impacto na segmentação.

Observando-se a Figura 8.2 pode-se constatar que o algoritmo falha na presença de lesões da SB, realçadas com os círculos vermelhos, segmentando-as como SC. Marcado com os círculos verdes encontram-se giros onde a SB é estreita, sendo substituída, de forma errada, por SC à medida que se aumenta o impacto do MRF, verificando-se que se pode remover detalhes mais finos, por influência da vizinhança.

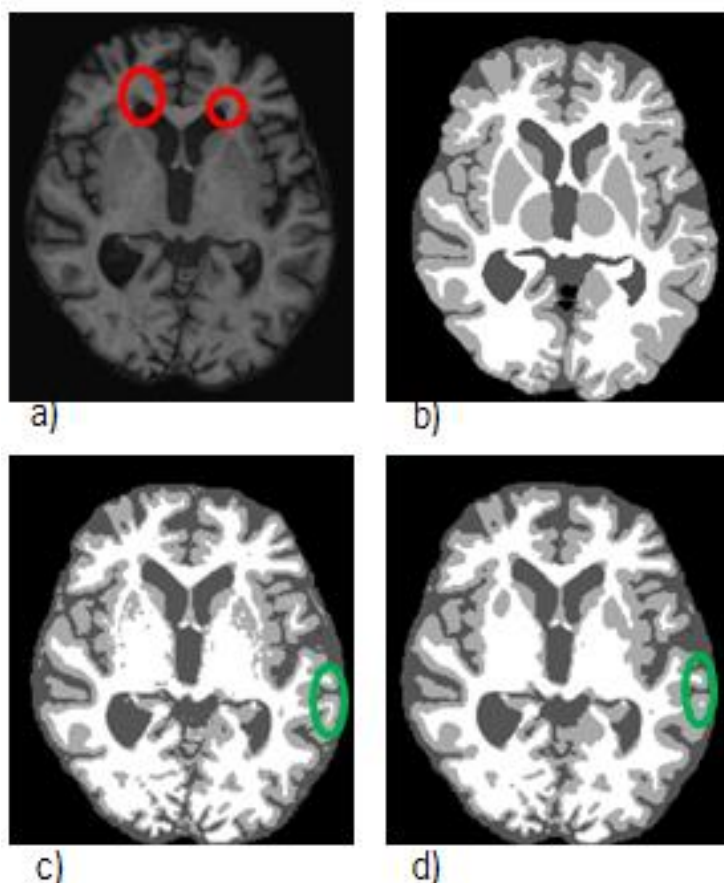


Figura 8.2 - *Slice* axial segmentado com o *Atropos*. Em a) observa-se a sequência T1, marcada com círculos vermelhos em locais de lesão da SB; em b) observa-se a marcação manual da imagem, enquanto em c) e d) se pode visualizar a segmentação do mesmo *slice*, mas com o parâmetro de suavização de 0,05 e 0,8, respectivamente.

8.4.2. Avaliação individual das características na classificação

Avaliou-se a utilização apenas das intensidades, e destas em conjunto com cada uma das outras características, individualmente, no algoritmo de classificação com RFs. Foram utilizadas 40 DTs, com 20 nós de profundidade máxima para todos os testes, uma vez que seria computacionalmente dispendioso fazer a procura dos melhores parâmetros para todos, para além de se ter considerado relevante manter as condições entre avaliações. Para que o treino fosse mais veloz, selecionou-se 100 mil pontos aleatórios de cada sujeito, sem estratificação,

para os quais se extraiu as características. Na Tabela 8.1 e Figura 8.3 encontram-se os resultados obtidos.

Tabela 8.1 - Avaliação de cada característica individualmente, em conjunto com as intensidades. Resultados para cada sujeito de treino, média e desvio padrão em coeficiente Dice. A - intensidades; B - intensidades e probabilidades *a posteriori*; C - intensidades e diferenças entre sequências, incluindo as estatísticas da vizinhança; D - intensidades e atlas; E - intensidades e módulo do gradiente.

		Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Média ± Desv. Pad.
A	LCR	0,679	0,671	0,705	0,660	0,656	0,674 ± 0,019
	SC	0,739	0,689	0,715	0,748	0,782	0,735 ± 0,035
	SB	0,778	0,635	0,699	0,760	0,797	0,734 ± 0,066
B	LCR	0,729	0,685	0,722	0,702	0,685	0,704 ± 0,020
	SC	0,762	0,782	0,800	0,836	0,851	0,806 ± 0,037
	SB	0,867	0,817	0,845	0,871	0,897	0,859 ± 0,030
C	LCR	0,721	0,700	0,717	0,707	0,682	0,705 ± 0,016
	SC	0,823	0,803	0,813	0,832	0,849	0,824 ± 0,018
	SB	0,861	0,833	0,871	0,872	0,907	0,869 ± 0,026
D	LCR	0,678	0,671	0,706	0,671	0,664	0,678 ± 0,016
	SC	0,764	0,749	0,744	0,771	0,806	0,767 ± 0,025
	SB	0,824	0,741	0,786	0,794	0,834	0,796 ± 0,037
E	LCR	0,682	0,680	0,712	0,669	0,662	0,681 ± 0,019
	SC	0,754	0,714	0,743	0,779	0,797	0,757 ± 0,032
	SB	0,794	0,660	0,740	0,781	0,803	0,755 ± 0,059

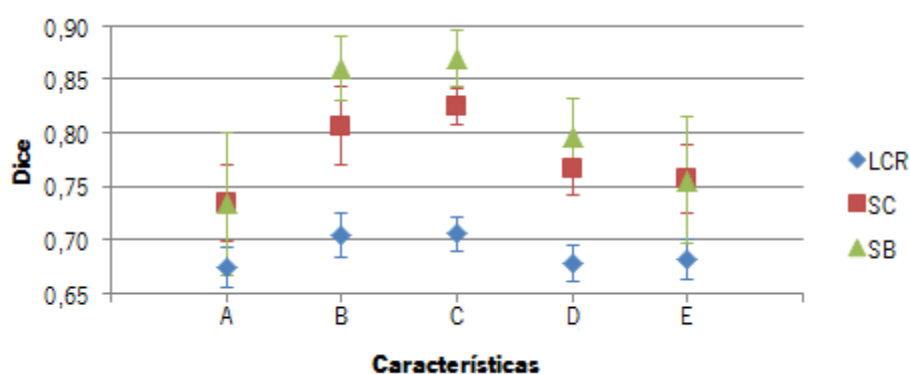


Figura 8.3 - Avaliação individual das características sobre a segmentação, quando utilizadas em conjunto com as intensidades das sequências. Resultados em média de Dice dos sujeitos de treino. A - intensidades; B - intensidades e probabilidades *a posteriori*; C - intensidades e diferenças entre sequências, incluindo as estatísticas da vizinhança; D - intensidades e atlas; E - intensidades e gradiente.

A utilização somente das intensidades (A na Tabela 8.1 e Figura 8.3) das sequências obtém os piores resultados, estando até abaixo do método do pacote *Atropos*. Isto deve-se, em grande parte, ao facto de as RFs não incorporarem informação da vizinhança, observando-se

muitos pontos isolados na SB na Figura 8.4(d). Mais, as lesões da SB, hiperintensidades em FLAIR (Figura 8.4(b)), foram mal classificadas como SC, enquanto alguns tecidos do exterior do cérebro foram tomados por SB. As intensidades, sozinhas, mostraram ser pouco robustas, resultando em performances da classificação variadas, refletindo-se no seu desvio padrão, que é elevado, em particular para a SB.

A inclusão das probabilidades *a posteriori* melhora os resultados anteriores (B na Tabela 8.1 e Figura 8.3), sendo consideravelmente superiores aos melhores do método *Atropos*, que se encontram na Figura 8.1. Comparativamente com a utilização das intensidades, todas as classes foram melhor classificadas, mas torna-se particularmente relevante na SB ao melhorar 0,125 em Dice, com redução para metade do desvio padrão anterior. Para isto contribuiu a incorporação indireta no modelo da RF da vizinhança dos *voxels*, com as probabilidades *a posteriori*, tornando a segmentação da referida região mais homogénea (Figura 8.4(e)). Verifica-se ainda que os tecidos externos à marcação manual, devido à remoção do crânio não ser ideal, são classificados como LCR, o que pode não estar correto, mas torna a segmentação mais credível, visualmente. Porém, uma porção de lesões da SB ainda é atribuída à SC.

Em C, na Tabela 8.1 e Figura 8.3, observa-se o resultado da junção das diferenças entre as sequências, das medidas estatísticas de vizinhança e suas diferenças entre regiões. Estas características são as que se traduzem na maior melhoria da segmentação, face às intensidades, não só porque em média é melhor, verificando-se o aumento de cerca de 0,09 e 0,135 para a SC e SB, respetivamente, relativamente a A, mas também é mais robusta, pois tem os desvios padrões mais baixos de todos. Para tal contribui a utilização das vizinhanças, à semelhança da característica anterior, mas as próprias diferenças também conseguem classificar corretamente as lesões, pois existe um desvio face às diferenças entre tecidos normais. No entanto, as referidas diferenças também podem explicar o aparecimento de SB entre os ventrículos (oval vermelho na Figura 8.4(f)), uma vez que, se se observar T1 e FLAIR, parece existir aí uma região estreita de intensidades de tecido normal, cuja diferença entre sequências pode ser semelhante à SB, e esteja reforçado devido ao número deste tipo de características.

Os gânglios da base são bem segmentados, tendo um aspeto homogéneo, apesar de se observar uma região de alguns *voxels* de SB interna mal classificada. O núcleo lentiforme do lado esquerdo quase não foi segmentado, talvez por ter pouco contraste, e o facto de ser estreito

tenha feito com que as diferenças fossem semelhantes à SB e seja ainda esbatido pelas medidas de vizinhança.

A utilização do atlas (D na Tabela 8.1 e Figura 8.3) traduz-se em resultados, em média, pouco superiores à utilização somente das intensidades, todavia o desvio padrão é inferior, sendo o mais reduzido, em conjunto com as diferenças, para o LCR. Para além disso, o desvio padrão da SC é o segundo mais baixo de todas as características. Isto pode ser o resultado de esta característica impor algum tipo de restrição espacial para ambas as classes, notando-se, na Figura 8.4(g) que existem menos pontos isolados do que na Figura 8.4(d).

Por fim, o gradiente (E na Tabela 8.1 e Figura 8.3) melhora um pouco a segmentação, relativamente à utilização apenas das intensidades, mas é a pior das características, com o segundo maior desvio padrão, o que indicia que a sua importância para as DTs é baixa (Figura 8.4(h)).

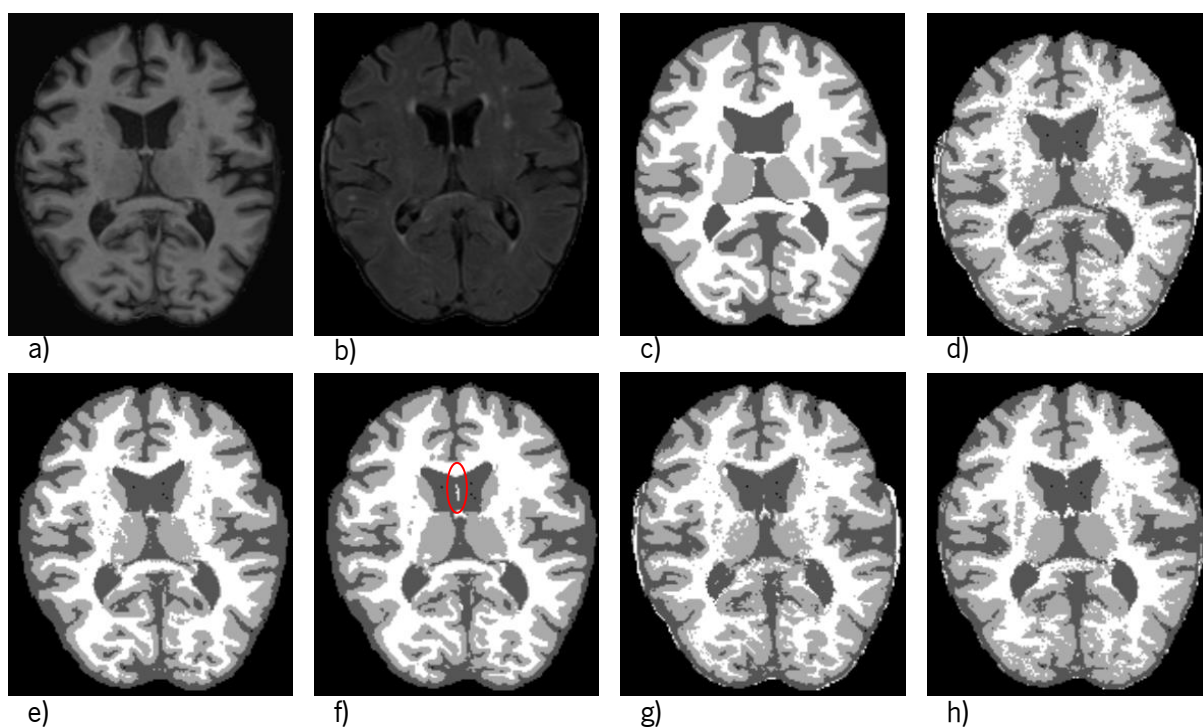


Figura 8.4 - Segmentação resultante de cada característica, em conjunto com as intensidades. a) T1, b) FLAIR, c) marcação manual, d) intensidades, e) intensidades e probabilidades *a posteriori*, f) intensidades, diferenças entre as sequências e medidas de vizinhança, g) intensidades e atlas, h) intensidades e gradiente.

8.4.3. Avaliação conjunta das características na classificação

Estudou-se o efeito da inclusão progressiva das características para a segmentação, utilizando 40 DTs com 20 nós de profundidade máxima, e extraindo 100 mil pontos aleatórios de cada sujeito de treino. Na Tabela 8.2 e Figura 8.5 encontram-se os resultados obtidos no

conjunto de treino, começando pelas intensidades (A) e adicionando progressivamente as probabilidades *a posteriori* (B), as diferenças entre sequências, medidas de vizinhança e suas diferenças (C), atlas (D) e gradiente E.

Tabela 8.2 - Avaliação conjunta das características. A - intensidades; B - intensidades e probabilidades *a posteriori*; C – intensidades, probabilidades *a posteriori*, diferenças entre sequências, medidas de vizinhança e suas diferenças; D - intensidades, probabilidades *a posteriori*, diferenças entre sequências, medidas de vizinhança e suas diferenças e atlas; E - intensidades, probabilidades *a posteriori*, diferenças entre sequências, medidas de vizinhança e suas diferenças, atlas e gradiente.

		Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Média ± Desv. Pad.
A	LCR	0,679	0,671	0,705	0,660	0,656	0,674 ± 0,019
	SC	0,739	0,689	0,715	0,748	0,782	0,735 ± 0,035
	SB	0,778	0,635	0,699	0,760	0,797	0,734 ± 0,066
B	LCR	0,729	0,685	0,722	0,702	0,685	0,704 ± 0,02
	SC	0,762	0,782	0,800	0,836	0,851	0,806 ± 0,037
	SB	0,867	0,817	0,845	0,871	0,897	0,859 ± 0,03
C	LCR	0,736	0,703	0,726	0,706	0,686	0,711 ± 0,02
	SC	0,831	0,814	0,812	0,843	0,856	0,831 ± 0,019
	SB	0,878	0,839	0,865	0,881	0,911	0,875 ± 0,026
D	LCR	0,734	0,707	0,735	0,718	0,690	0,717 ± 0,019
	SC	0,842	0,835	0,829	0,718	0,865	0,845 ± 0,015
	SB	0,891	0,850	0,875	0,886	0,916	0,883 ± 0,024
E	LCR	0,738	0,708	0,735	0,717	0,690	0,718 ± 0,02
	SC	0,845	0,837	0,828	0,857	0,866	0,847 ± 0,015
	SB	0,891	0,853	0,873	0,887	0,916	0,884 ± 0,023

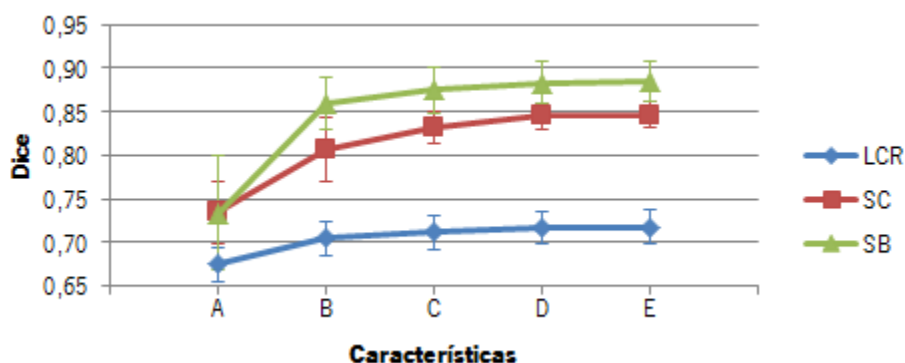


Figura 8.5 - Avaliação conjunta das características, por classe, em média. A - intensidades; B - intensidades e probabilidades *a posteriori*; C – intensidades, probabilidades *a posteriori*, diferenças entre sequências, medidas de vizinhança e suas diferenças; D - intensidades, probabilidades *a posteriori*, diferenças entre sequências, medidas de vizinhança e suas diferenças e atlas; E - intensidades, probabilidades *a posteriori*, diferenças entre sequências, medidas de vizinhança e suas diferenças, atlas e gradiente.

Numa primeira análise, observa-se que todas as características contribuem para o aumento do Dice, em média, existindo algumas cujo impacto foi superior. Esta melhoria é

sobretudo notória para a SB e SC. O LCR melhora menos, mas isso deve-se ao método de remoção do crânio, como se discutirá mais à frente.

Observando os resultados, verifica-se que as probabilidades *a posteriori* melhoram significativamente a média, em Dice, como se discutiu na secção anterior, principalmente a SB; o sujeito com pior resultado neste caso é superior ao melhor da utilização apenas das intensidades. A inclusão das diferenças e das medidas estatísticas da vizinhança mostrou ser muito relevante, em particular na SC, ao aumentar o Dice médio dessa classe em 0,025, mas também porque diminuiu o desvio padrão desse tecido para cerca de metade, o que torna a sua segmentação mais robusta. Para isto terá contribuído a remoção de pontos isolados de SC na SB, ou seja FPs, que ainda restavam com a inclusão das probabilidades *a posteriori* (Figura 8.4), e a correta classificação das lesões da SB. Estes fatores diminuem também o número de FN e aumentam os VP da SB, contribuindo para a sua melhoria de cerca de 0,016.

Por seu lado, a contribuição do atlas foi menor, como se previa, mas ainda assim positiva, traduzindo-se na melhoria de 0,014 e 0,008 para a SC e SB, respetivamente. Aliado a isto, o desvio padrão diminuiu em todas as classes, tornando as segmentações mais coerentes entre sujeitos.

O módulo do gradiente apenas se refletiu na terceira casa decimal das médias de todas as classes. No entanto, optou-se por manter esta característica, uma vez que, ainda assim, teve uma contribuição positiva; devendo-se, no futuro, utilizar um método mais objetivo de verificação da importância das características, que forneça dados mais concretos acerca da sua inclusão ou remoção.

Finalmente, medindo as diferenças das médias entre todas as características e a utilização apenas das intensidades verifica-se que houve melhorias de 0,044, 0,112 e 0,15, em Dice, para as classes LCR, SC e SB, respetivamente.

8.4.4. Avaliação da remoção de ruído

Testou-se o impacto que a remoção do ruído teria na segmentação, utilizando-se florestas com os mesmos parâmetros da secção anterior. Os resultados podem ser consultados na Tabela 8.3.

Tabela 8.3 - Avaliação da remoção do ruído na segmentação.

	Sem remoção de ruído	Com remoção de ruído
LCR	0,718 ± 0,020	0,720 ± 0,018
SC	0,847 ± 0,015	0,846 ± 0,018
SB	0,884 ± 0,23	0,881 ± 0,024

Observando a Tabela 8.3, verifica-se que a remoção do ruído apenas beneficiou a classe de LCR. Relativamente à SC e SB, contribuiu negativamente, reduzindo a média e aumentando o desvio padrão. Estes resultados explicam-se com a aquisição das RM, uma vez que a utilização de um campo magnético de 3 T e a distância entre *slices* ser de 3 mm contribuem para a melhor relação sinal-ruído das imagens. Assim, se a quantidade deste artefacto era reduzida, a aplicação do método para a sua remoção pode ter apenas contribuído para esbater alguns detalhes, nomeadamente nas fronteiras entre a SC e SB. Portanto, para o treino das RFs optou-se por não remover o ruído, porém esse problema deve ser tratado quando se utiliza as sequências MPRAGE para os EVR.

8.4.5. Avaliação do número de pontos e classes de treino

Utilizando todas as características, avaliou-se a influência do número de pontos e de classes utilizados no treino. Assim, testou-se o efeito da utilização de todos os pontos disponíveis para treino, mantendo as classes anteriores; posteriormente, avaliou-se a utilização de todos os pontos e 6 classes de treino: SC cortical, gânglios da base, SB, lesões da SB, ventrículos e LCR do exterior do encéfalo. Para o cálculo dos resultados, fundiu-se as 6 *labels* nas 3 classes utilizadas para validação. Os resultados encontram-se na Tabela 8.4, para uma RF com 40 DTs e profundidade limite de 20 nós, enquanto na Figura 8.6 se pode observar exemplos de segmentação.

Tabela 8.4 - Avaliação do número de pontos e classes para treinar. Os resultados representam a média e desvio padrão no coeficiente de Dice para cada classe, e o tempo de treino por avaliação corresponde ao tempo gasto apenas no treino da RF, para os 5 treinos de cada procedimento *leave-one-out*.

	100 mil pontos p/ sujeito e 3 classes	Todos os pontos e 3 classes	Todos os pontos e 6 classes
LCR	0,718 ± 0,02	0,718 ± 0,02	0,716 ± 0,018
SC	0,847 ± 0,015	0,848 ± 0,015	0,847 ± 0,015
SB	0,884 ± 0,23	0,886 ± 0,023	0,886 ± 0,023
Tempo de treino por avaliação (seg)	215	1315	1405

Analisando-se os resultados, constata-se que a utilização de todos os pontos de treino traduz-se numa melhoria ligeira da média, apenas na terceira casa decimal, mantendo-se o desvio padrão, relativamente à utilização de 100 mil pontos aleatórios de cada sujeito.

Já a utilização de 6 classes de treino, não se traduziu em nenhuma melhoria, descendo a média em duas *labels*, comparativamente à utilização de 3 classes e todos os pontos de treino.

Para além disso, observando a Figura 8.6(d) constata-se que, apesar de em média os resultados serem semelhantes, uma parte dos gânglios da base, particularmente o núcleo caudado (elipses vermelhas) são classificados como SC do córtex, ou seja, os resultados só são semelhantes porque há fusão das classes. Isto evidencia a capacidade de generalização das RFs, que utilizando um treino mais geral, com 3 classes, consegue construir um modelo tão, ou mais, preciso do que utilizando uma abordagem mais específica com mais categorias. Assim sendo, optou-se por manter a abordagem inicial nas futuras avaliações, isto é, treinar com as 3 classes mais gerais.

Mediu-se o tempo gasto no treino por cada avaliação *leave-one-out* (Tabela 8.4), ou seja, 5 treinos, utilizando um *laptop* com processador *Intel Core i7-3610QM* a 2,3 GHz e 6 GB de memória RAM. Assim, verificou-se que a utilização de todos os pontos (1.837.728 pontos) implica despende mais de 6 vezes o tempo necessário quando se utilizam 100 mil pontos por sujeito (400.000 pontos). Assim sendo, optou-se por continuar a utilizar a seleção aleatória de 100 mil pontos, uma vez que o impacto não é muito significativo nos resultados, e a carga computacional diminui, permitindo realizar-se mais testes.

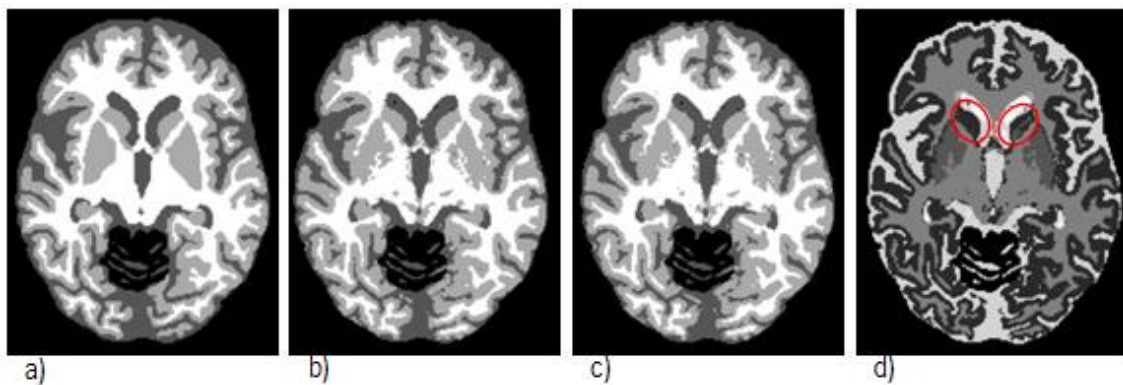


Figura 8.6 - Exemplos de segmentação, variando o número de pontos e de classes de treino. Em a) encontra-se a marcação manual, em b) a segmentação com 3 classes e 100 mil pontos de treino, em c) a segmentação com 3 classes e todos os pontos de treino e em d) a segmentação de 6 classes e todos os pontos de treino.

8.4.6. Avaliação do critério de divisão dos nós

Avaliou-se o critério de decisão de divisão dos nós, testando-se o coeficiente Gini, que foi utilizado nos testes anterior, e o ganho de informação. Os resultados encontram-se na Tabela 8.5.

Tabela 8.5 - Avaliação do critério de divisão nos nós. Resultados mostram a média da segmentação e desvio padrão, em Dice.

	Coeficiente Gini	Ganho de informação
LCR	0,718 ± 0,02	0,718 ± 0,02
SC	0,847 ± 0,015	0,846 ± 0,014
SB	0,884 ± 0,23	0,884 ± 0,023

Analisando os resultados, verifica-se que apenas existem diferenças na SC, mas estas são muito reduzidas. Assim se pode concluir que analisar o aumento da informação é semelhante à redução da taxa de erro esperada, sendo ambas medidas de pureza dos nós, no contexto do problema de segmentação de tecidos, e com as características escolhidas.

Como o coeficiente de Gini demonstrou ser equivalente ao ganho de informação e é a opção por defeito da RF, optou-se por manter esse critério nos testes seguintes.

8.4.7. Avaliação do número de *Decision Trees* e profundidade máxima dos nós

Com os parâmetros escolhidos anteriormente, isto é, escolha aleatória de 100 mil pontos de treino por sujeito e coeficiente de Gini como critério de avaliação da divisão dos nós, testou-se a influência do número de DTs e profundidade máxima dos nós na segmentação. Os resultados para cada combinação e classe podem ser consultados na Tabela 8.6, enquanto as médias do conjunto do LCR, SC e SB se encontram traçadas na Figura 8.7.

Analisando os resultados, verifica-se que a diferença entre a melhor combinação de parâmetros e a pior não é superior a um desvio padrão; além disso, a diferença nunca é superior a 0,01 em nenhuma classe. Assim, a variação das combinações não tem um forte impacto na segmentação do conjunto de treino, mas pode ter na generalização para o conjunto de teste.

Em primeira análise, observando-se a Figura 8.7 constata-se que a utilização de 10 nós como profundidade máxima permitida foi a pior opção para todas as RFs, o que se pode dever ao facto de existirem 87 características e essa profundidade não ser suficiente para que o grau de certeza seja mais elevado, em cada DT. O aumento da profundidade máxima para 15 foi aquele que teve a maior taxa de melhoria para todas as florestas testadas, indicando que esse é o mínimo aceitável, permitindo que as DTs já consigam escolher um conjunto de características que lhes permita classificar com maior certeza.

Observando a RF com 20 árvores, aumentar a profundidade máxima a partir de 15 nós traduziu-se na deterioração progressiva da qualidade da segmentação, particularmente para o número de nós mais elevado. Aqui verificou-se um problema de *overfitting*, tendo as RFs ficado

muito longas e demasiado concentradas nos exemplos de treino, e não existindo mais DTs que, com a sua aleatoriedade, pudessem compensar o problema. As florestas de 30 árvores estabilizam a partir dos 15 nós, inclusive, até que os resultados pioram para 30 de profundidade, indicando que até aí o número de árvores tinha suprimido o *overfitting*, mas não foi suficiente a partir desse ponto.

Tabela 8.6 - Avaliação do número de *Decision Trees* e da profundidade máxima dos nós, para cada classe. Resultados da segmentação em Dice.

	Nós	Número de DTs				
		20	30	40	50	60
LCR	10	0,715 ± 0,018	0,714 ± 0,017	0,714 ± 0,019	0,713 ± 0,018	0,713 ± 0,016
	15	0,716 ± 0,019	0,717 ± 0,019	0,716 ± 0,018	0,716 ± 0,019	0,717 ± 0,018
	20	0,715 ± 0,020	0,716 ± 0,020	0,717 ± 0,019	0,717 ± 0,019	0,717 ± 0,020
	25	0,716 ± 0,021	0,717 ± 0,019	0,717 ± 0,020	0,717 ± 0,020	0,718 ± 0,020
	30	0,715 ± 0,020	0,717 ± 0,020	0,717 ± 0,020	0,717 ± 0,020	0,718 ± 0,020
SC	10	0,840 ± 0,015	0,838 ± 0,018	0,838 ± 0,017	0,838 ± 0,015	0,839 ± 0,017
	15	0,844 ± 0,016	0,845 ± 0,016	0,845 ± 0,015	0,845 ± 0,015	0,847 ± 0,015
	20	0,844 ± 0,014	0,845 ± 0,014	0,846 ± 0,015	0,847 ± 0,014	0,847 ± 0,014
	25	0,842 ± 0,014	0,845 ± 0,015	0,846 ± 0,014	0,846 ± 0,015	0,846 ± 0,014
	30	0,842 ± 0,014	0,844 ± 0,014	0,845 ± 0,015	0,846 ± 0,014	0,846 ± 0,014
SB	10	0,878 ± 0,024	0,877 ± 0,025	0,8778 ± 0,025	0,878 ± 0,025	0,878 ± 0,025
	15	0,882 ± 0,024	0,883 ± 0,024	0,883 ± 0,024	0,884 ± 0,023	0,884 ± 0,023
	20	0,883 ± 0,023	0,884 ± 0,023	0,885 ± 0,023	0,885 ± 0,023	0,885 ± 0,023
	25	0,883 ± 0,023	0,884 ± 0,023	0,884 ± 0,023	0,885 ± 0,023	0,885 ± 0,022
	30	0,882 ± 0,023	0,883 ± 0,023	0,884 ± 0,023	0,884 ± 0,023	0,885 ± 0,023

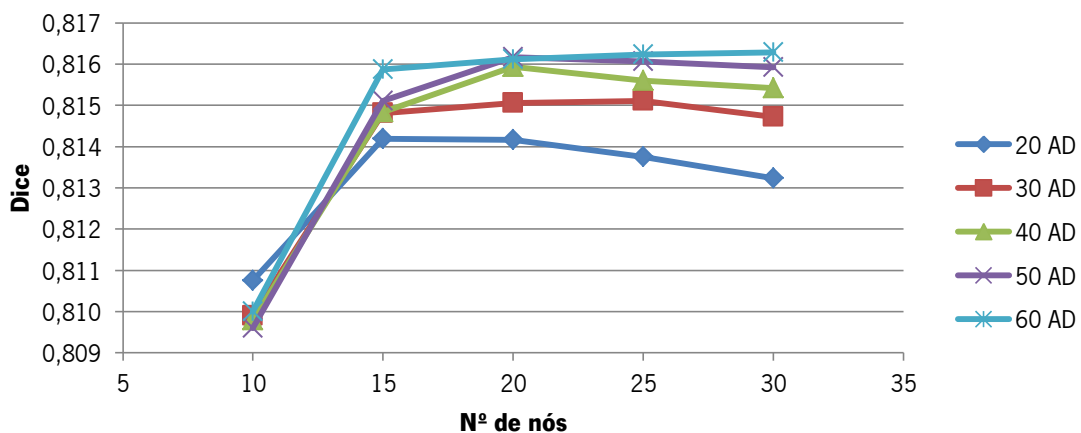


Figura 8.7 - Avaliação do número de *Decision Trees* e profundidade máxima, para o conjunto das três classes.

As florestas com 40 e 50 DTs mostraram comportamentos semelhantes, atingindo o seu máximo nos 20 nós, e piorando a partir daí. Todavia, a maior das duas demonstrou ser mais robusta, com taxas de deteriorações menos acentuadas. Já a RF com 60 árvores não sofreu *overfitting* na gama de valores de profundidades máximas testadas, tendo melhorado sempre os seus resultados, ainda que a taxas reduzidas a partir dos 20 nós.

Assim se comprova que o aumento do número de DTs tem impacto positivo, ainda que a partir de dada altura esse impacto seja reduzido. Já o mesmo não se pode dizer para número de nós máximo, verificando-se que pode ter influência negativa, se for demasiado elevado.

A escolha de 40 DTs e 20 nós de profundidade máxima para as experiências das secções anteriores revelou-se uma escolha acertada, já que nesse ponto se verifica o máximo das RFs com 40 classificadores fracos, estando próximos dos melhores resultados obtidos entre todas as outras RFs. Porém, essas segmentações não foram ótimas.

Para a escolha dos melhores parâmetros analisou-se a Tabela 8.6, procurando-se pelos resultados máximos de cada classe, em média, e entre esses os que tinham desvio padrão mais reduzido, realçando-se as combinações que preenchessem os requisitos. A maior parte dos resultados assim encontrados deram-se em RFs com 60 DTs, coincidindo no LCR e SB para a profundidade máxima de 25 nós, que no caso da SC não é o máximo, mas é muito próximo. O aumento do número de DTs para mais de 60, provavelmente, traduzir-se-ia na melhoria dos resultados; no entanto pode-se verificar que entre as florestas com 50 e 60 árvores as diferenças são inferiores à terceira casa decimal, pelo que se poderia esperar que esse aumento se refletisse em diferenças ainda menores. Assim optou-se por uma floresta com 60 DTs. Também se decidiu pela imposição de 25 nós como profundidade máxima por ser o máximo para duas das classes mas, adicionalmente, porque nas florestas com 40 e 50 árvores o ponto a partir do qual parece ter-se começado a sentir o efeito do *overfitting* foi quando existiam mais nós do que a metade do número de DTs.

8.4.8. Avaliação do efeito do pós-processamento

Após a avaliação dos parâmetros do classificador, avaliou-se o efeito da remoção de pequenas regiões isoladas. Assim, considerou-se que se poderiam remover regiões isoladas que fossem constituídas por menor número, ou igual, do que 4 *voxels*, uma vez que se se aumentar este valor podem existir pequenas regiões bem classificadas que podem ser trocadas, nomeadamente em sulcos estreitos, ou mesmo SB em giros constituídos por muita SC. Na Tabela 8.7 pode-se consultar os resultados.

Tabela 8.7 - Avaliação do pós-processamento para a correção de pequenas regiões isoladas, testando-se o número de pontos máximos de cada região isolada.

	0 pontos	1 ponto	2 pontos	3 pontos	4 pontos
LCR	0,718 ± 0,020	0,719 ± 0,019	0,718 ± 0,019	0,718 ± 0,019	0,718 ± 0,019
SC	0,846 ± 0,014	0,847 ± 0,014	0,847 ± 0,014	0,848 ± 0,014	0,848 ± 0,014
SB	0,885 ± 0,022	0,885 ± 0,023	0,885 ± 0,023	0,885 ± 0,023	0,885 ± 0,023

Assim se verifica que o procedimento tem um impacto reduzido. Na SB nunca altera a média da segmentação, mas aumenta ligeiramente o desvio padrão. Por seu lado, na SC a média melhora até duas unidades da terceira casa decimal, mantendo-se o desvio padrão, enquanto no LCR apenas a remoção de um máximo de um ponto isolado é que tem efeito na média, porém o desvio padrão diminui se se aumentar o número de pontos a remover.

Na Figura 8.8 observa-se um *slice* axial antes e depois do pós-processamento. Verifica-se que o procedimento para corrigir os pontos do LCR com intensidades negativas, que ficaram de fora da classificação é eficaz, notando-se o seu efeito, principalmente, nos ventrículos. Relativamente ao pós-processamento para remoção de pequenas regiões de tecidos, o seu efeito visual é dificilmente detetado. No entanto, quando se substituem regiões iguais ou inferiores a três *voxels*, também se alteram algumas que estariam corretamente segmentadas, como se encontra marcada com os círculos vermelhos na Figura 8.8, o que não se verifica se esse *threshold* for de um *voxel*. Assim sendo, apesar de o máximo de melhoria na SC ter sido para zonas com três elementos, optou-se por aplicar este procedimento do pós-processamento para regiões inferiores ou iguais a um elemento, sendo que esta melhora uma unidade na terceira casa decimal da SC, mas também do LCR.

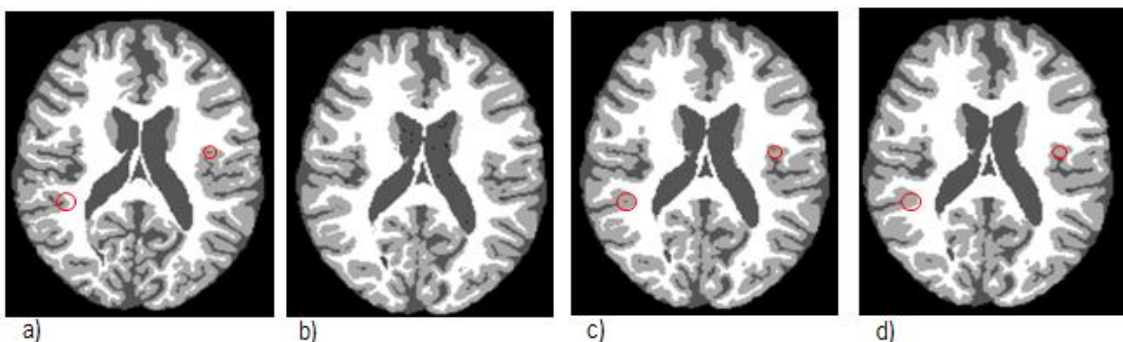


Figura 8.8 - Efeito do pós-processamento. Em a) encontra-se as marcações manuais, em b) a imagem segmentada sem pós-processamento, em c) a imagem segmentada com pós-processamento e remoção de regiões com um ponto e em d) a imagem segmentada com pós-processamento e remoção de regiões até 3 *voxels*.

8.4.9. Avaliação no conjunto de teste

Aplicou-se o algoritmo de segmentação no conjunto de teste, e enviou-se as imagens para o MRBrainS para avaliação; os resultados encontram-se na Tabela 8.8.

Tabela 8.8 - Resultados da segmentação no conjunto de teste.

	Dice	Dist. Haus. Mod. (mm)	Dif. Abs. Vol. (%)
LCR	0,683 ± 0,039	4,93 ± 2,48	34,13 ± 24,70
SC	0,834 ± 0,026	2,85 ± 0,91	9,84 ± 5,90
SB	0,876 ± 0,017	2,69 ± 0,51	9,16 ± 5,63

Analisando-se a Tabela 8.8, e comparando com os resultados após a aplicação do pós-processamento no conjunto de treino para remoção de regiões isoladas de um *voxel* da Tabela 8.7, verifica-se que todas as classes obtiveram resultados inferiores aos reportados no conjunto de treino.

O LCR foi a classe cujos resultados foram mais baixos, tal como se verificava nas imagens de treino, tendo diminuído cerca de 0,036 na média do Dice relativamente aos testes anteriores, e aumentado o desvio padrão.

O maior problema do LCR é o processo de remoção do crânio, que não elimina algumas regiões na base do mesmo, junto ao cerebelo, que acabam por ser classificadas como líquido. Na Figura 8.9 observa-se a sequência T1, a marcação manual e a segmentação, evidenciando-se nesta última com um círculo vermelho a região do cerebelo e tronco cerebral que são removidos, aquando da avaliação, não contribuindo para os resultados. Fora do referido círculo, observam-se regiões classificadas maioritariamente como LCR que não foram excluídas pelo processo de remoção do crânio, mas que não constam do *ground truth*, sendo FPs. Assim, dado essas regiões serem extensas, se explica os resultados muito inferiores do LCR, relativamente às outras classes, e a elevada percentagem da diferença absoluta de volume. Também se explica os desvios padrões elevados, estando estes associados à remoção do crânio, que varia um pouco de sujeito para sujeito.

A performance da SC e da SB diminuíram, em Dice, 0,013 e 0,009, respetivamente, relativamente às avaliações com os dados de treino. Esta variação não é muito acentuada, verificando-se que o classificador foi robusto na presença de novos dados, mesmo só existindo cinco sujeitos de treino, o que corresponde a um quarto de toda a base de dados.

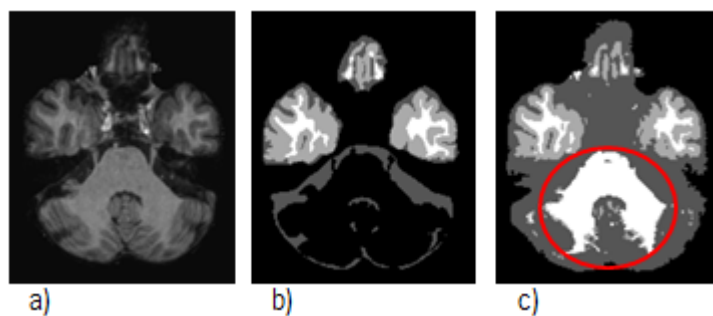


Figura 8.9 - Impacto da remoção do crânio sobre o líquido cefalorraquidiano. Da esquerda para a direita observa-se a sequência T1 depois de removido o crânio, a marcação manual e a segmentação obtida. Com o círculo vermelho evidencia-se a região do cerebelo e tronco cerebral.

A distância de Hausdorff da SC e da SB são consideradas positivas, uma vez que, em média, os FP se encontram a menos de 3 mm da verdadeira segmentação. A melhor remoção do crânio permitiria melhorar esta métrica, pois na Figura 8.9 observam-se alguns FPs destas classes nas regiões que deveriam ser excluídas, ou seja, se essas tivessem sido removidas, o 95º percentil das distâncias de Hausdorff estaria em medidas mais reduzidas.

A diferença absoluta de volume não permite concluir se o algoritmo falha por excesso ou por defeito. Para esta métrica o LCR apresenta valores muito elevados, devido maioritariamente aos problemas supramencionados, mas também existem sulcos onde por vezes é classificado como SC. Para a SC e SB este já é inferior a 10%, porém deve-se reduzir o desvio padrão, no futuro.

Na Figura 8.10 observa-se as sequências T1 e FLAIR, e a segmentação de um sujeito do conjunto de teste cujos resultados foram próximos da média. Verifica-se que a segmentação foi bem-sucedida, mesmo nas lesões da SB e em regiões hiperintensas dos ventrículos.

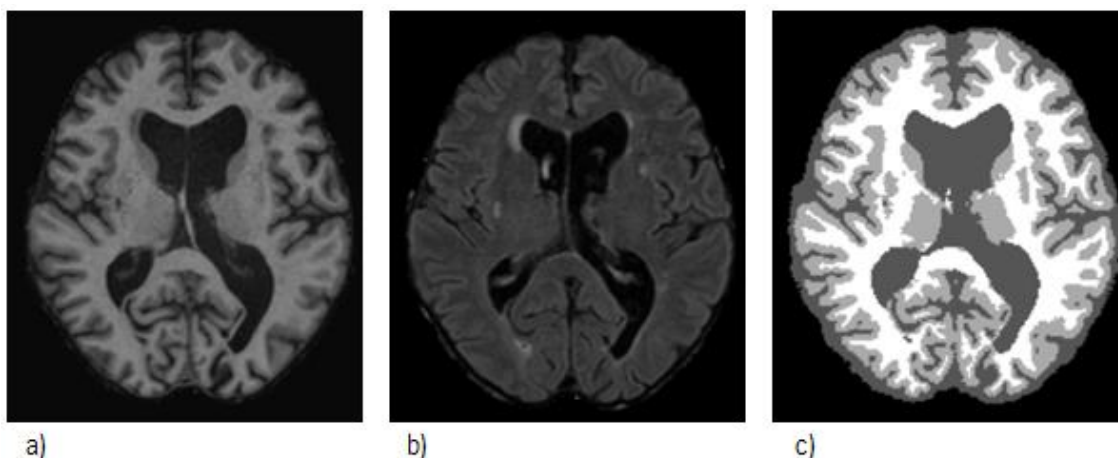


Figura 8.10 - Segmentação de um sujeito do conjunto de teste. a) T1; b) FLAIR; c) segmentação.

8.4.10. Avaliação da segmentação utilizando apenas a sequência T1

Após a validação do algoritmo de segmentação de tecidos, avaliou-se a sua performance se apenas existisse a sequência T1, de forma a testar a sua aplicabilidade na extração de SB para os EVR. Neste caso, as características que calculam as diferenças entre sequências deixam de existir, o que reduz o número total de características para 20. Assim sendo, considerou-se que não deveriam existir mais nós do que características, apesar de a escolha destas nos nós ser aleatória, pelo que se colocou o limite máximo do teste nesse valor, escolhendo-se o limite de 40 DTs para a avaliação. Os resultados podem ser consultados na Tabela 8.9 e Figura 8.11.

Tabela 8.9 - Avaliação do número de *Decision Trees* e da profundidade máxima dos nós, para cada classe. Resultados da segmentação em Dice, utilizando apenas a sequência T1 de base.

	Nós	Nº de DTs			
		10	20	30	40
LCR	5	0,705 ± 0,021	0,706 ± 0,014	0,706 ± 0,016	0,705 ± 0,013
	10	0,712 ± 0,018	0,712 ± 0,018	0,711 ± 0,019	0,712 ± 0,018
	13	0,712 ± 0,019	0,712 ± 0,019	0,712 ± 0,018	0,713 ± 0,018
	15	0,713 ± 0,021	0,713 ± 0,020	0,713 ± 0,020	0,712 ± 0,019
	17	0,712 ± 0,020	0,713 ± 0,020	0,71 ± 0,020	0,713 ± 0,019
	20	0,711 ± 0,021	0,712 ± 0,020	0,712 ± 0,020	0,713 ± 0,019
SC	5	0,814 ± 0,019	0,816 ± 0,021	0,813 ± 0,024	0,814 ± 0,021
	10	0,832 ± 0,017	0,834 ± 0,017	0,834 ± 0,015	0,834 ± 0,017
	13	0,835 ± 0,016	0,835 ± 0,017	0,836 ± 0,017	0,836 ± 0,017
	15	0,835 ± 0,016	0,836 ± 0,017	0,836 ± 0,016	0,836 ± 0,017
	17	0,833 ± 0,018	0,836 ± 0,017	0,835 ± 0,016	0,837 ± 0,017
	20	0,832 ± 0,017	0,835 ± 0,017	0,835 ± 0,018	0,836 ± 0,017
SB	5	0,862 ± 0,034	0,861 ± 0,035	0,861 ± 0,035	0,861 ± 0,034
	10	0,872 ± 0,029	0,874 ± 0,029	0,874 ± 0,029	0,874 ± 0,030
	13	0,875 ± 0,028	0,876 ± 0,028	0,877 ± 0,027	0,877 ± 0,028
	15	0,875 ± 0,026	0,877 ± 0,027	0,877 ± 0,027	0,878 ± 0,027
	17	0,874 ± 0,027	0,874 ± 0,027	0,876 ± 0,027	0,878 ± 0,026
	20	0,872 ± 0,027	0,876 ± 0,026	0,877 ± 0,026	0,878 ± 0,026

Observando a Figura 8.11, verifica-se que o aumento da profundidade máxima permitida de 5 para 10 nós conduz à melhoria do resultados em cerca de 0,012, o que demonstra que 5 nós não são suficientes para que a decisão das árvores tenha a melhor performance na segmentação.

Constatou-se, ainda, que a RF com 10 árvores foi, na globalidade, a que demonstrou pior performance. A partir da estipulação do número máximo de nós acima de 13, os seus

resultados deterioraram-se, em consequência do *overfitting* em relação ao treino, uma vez que o número de DTs não foi suficiente para colmatar esse problema.

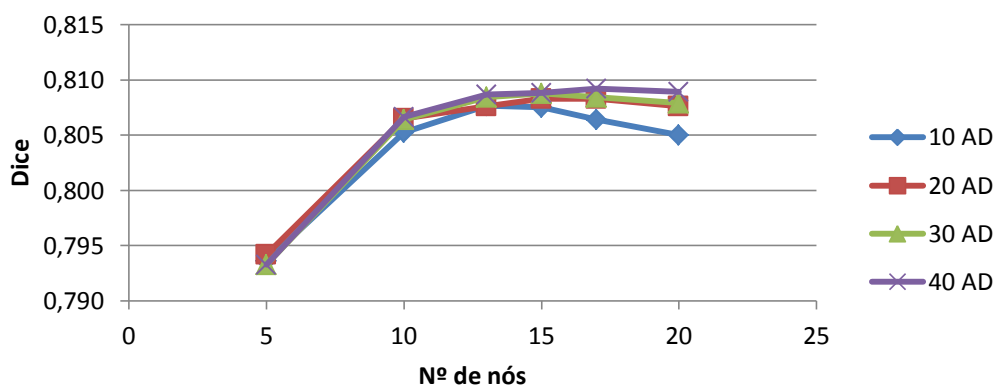


Figura 8.11 - Avaliação do número de *Decision Trees* e da profundidade máxima, para o conjunto das três classes, utilizando apenas a sequência T1 como base.

Tanto as RFs com 20 árvores como as constituídas por 30 começaram a sentir o efeito do aumento excessivo da profundidade máxima permitida a partir dos 15 nós, ao verem os seus resultados diminuir. No entanto, a taxa de redução foi inferior à floresta anterior, demonstrando o poder que o número de árvores tem sobre o *overfitting*.

Por seu lado, a floresta com 40 DTs foi mais robusta, já que o número de classificadores fracos foi o mais elevado, suprimindo o efeito de *overfitting* através de mais árvores com treinos diferentes. No entanto, entre 13 e 15 nós os seus resultados estabilizaram, sofrendo um aumento quando se permitiu que a profundidade máxima fosse 17 nós, não estando de acordo com o que seria esperado tendo em conta as outras florestas.

Analisando-se a Tabela 8.9 verifica-se que o máximo de cinco nós foi negativo para todas as classes, particularmente para a SC, que registou aumentos superiores a 0,02 a partir dos 10 nós, indicando que esta é a classe para a qual as RFs necessitaram de mais informação para classificar, ou seja, é a mais incerta. O motivo para esse comportamento pode ser o facto de nos seus sulcos estar em contacto com SB e LCR, sendo regiões onde em alguns casos há zonas das referidas classes muito estreitas, e sujeitas ao efeito de volume parcial, sendo classificados com essas classes.

Ainda na Tabela 8.9, realçou-se, para cada classe, os resultados máximos. Assim, optou-se por utilizar a RF com 40 DTs e 17 nós de profundidade, por ter sido aquela com resultados máximos para SC e SB. Aqui colocam-se algumas reservas, porque 17 nós está próximo do número de características disponíveis (20) quando se tem apenas a sequência T1, o que pode

conduzir a um problema de *overfitting*, apesar de os resultados da sua segmentação terem sido os melhores. Porém, espera-se que ao utilizar 40 DTs, esse problema, se existir, seja reduzido.

Na Figura 8.12 observa-se um exemplo de segmentação. Utilizar características apenas da sequência T1 fez com que o classificador não conseguisse reconhecer as lesões da SB (Figura 8.12(e)), mesmo quando a sua área era considerável, pois as intensidades em T1 (Figura 8.12(a)) são semelhantes às da SC. Na Figura 8.12(f) encontra-se, para efeito de comparação, o mesmo *slice* classificado com a extração de características de todas as sequências, verificando-se que as lesões foram corretamente segmentadas como SB, ou seja, a utilização de mais sequências, provavelmente a FLAIR, é útil para esses casos.

Comparando com os melhores resultados da Tabela 8.6, a utilização apenas de T1 para as características obteve resultados inferiores, ou iguais, a 0,01, devendo-se aos FP da SC e FN da SB. Já em relação ao método *Atropos* (Figura 8.1), o algoritmo proposto obteve resultados superiores em todas as classes, registrando melhorias superiores a 0,1 para a SC e a 0,06 na SB.

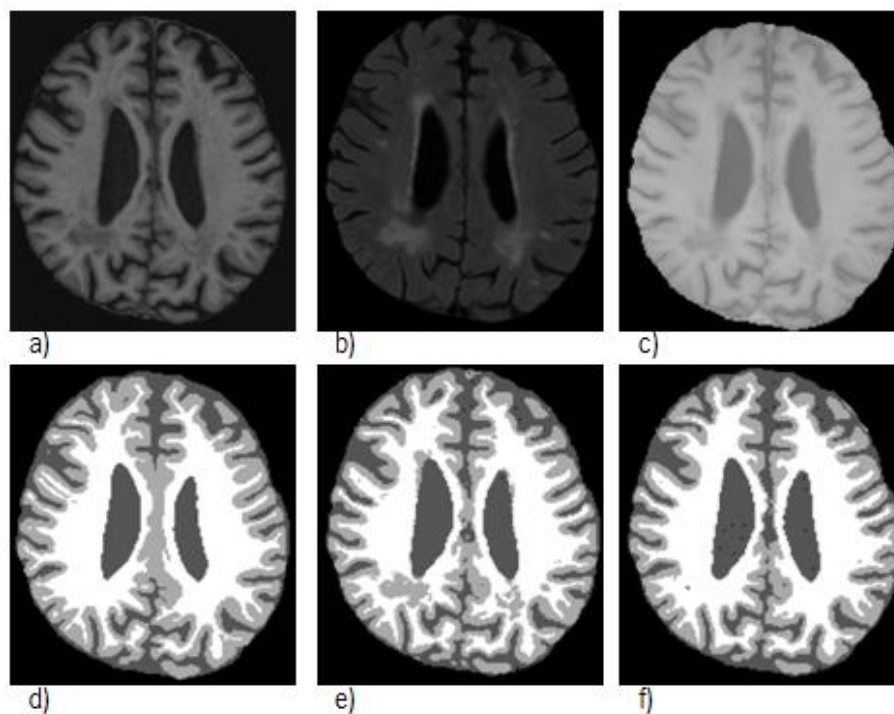


Figura 8.12 - Exemplo de segmentação utilizando apenas T1. a) T1; b) FLAIR; c) T1 IR; d) segmentação manual; e) segmentação com características apenas da sequência T1; f) segmentação extraíndo características de todas as sequências.

8.4.11. Avaliação da segmentação de sequências MPRAGE

Após a validação da segmentação utilizando apenas a sequência T1, testou-se a sua aplicabilidade nas imagens de sequência MPRAGE utilizadas na segmentação de EVR. Na Figura 8.13 encontram-se as segmentações de dois sujeitos.

A segmentação foi bem-sucedida, mesmo nos gânglios da base e utilizando extração de características em 2D. Porém, as lesões dos sujeitos foram mal classificadas como SC ou LCR, evidenciando a necessidade da utilização da sequência FLAIR.

O cerebelo não foi corretamente segmentado, sendo maioritariamente associado com a SC, todavia este problema foi minimizado ao utilizar a segmentação do *Freesurfer*, numa fase posterior para a remoção dessa região.

Assim, utilizou-se a segmentação obtida com o algoritmo proposto, quer para a extração de SB, quer para o cálculo das estatísticas dos tecidos no algoritmo de deteção dos EVR.

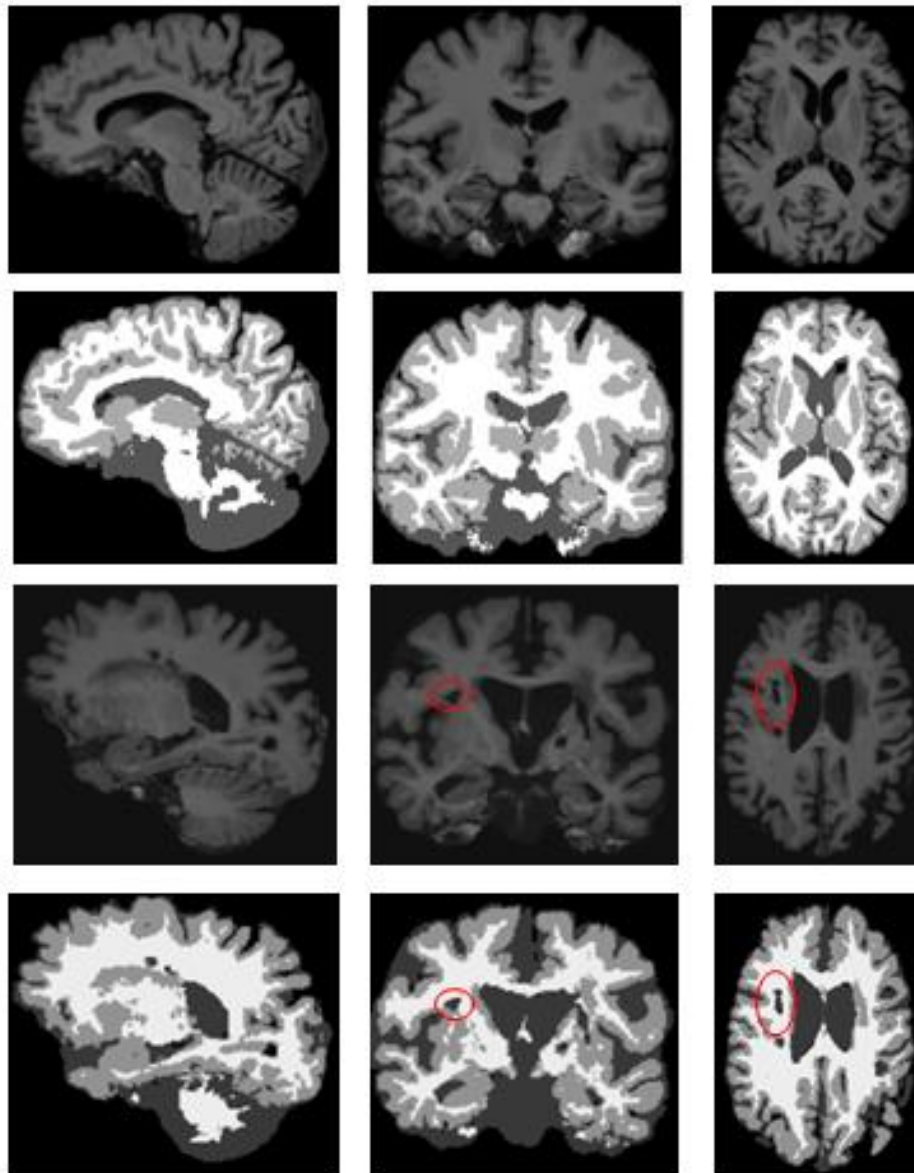


Figura 8.13 - Segmentação de tecidos em sequências MPRAGE. Na primeira e terceira linha observam-se as RM de dois sujeitos, enquanto na segunda e quarta linha se encontram as segmentações obtidas, respectivamente. Nas colunas, da esquerda para a direita são os *slices* sagital, coronal e axial. Nos círculos a vermelho marcam-se lesões.

8.5. Detecção automática de Espaços de Virchow-Robin

Na deteção de EVR utilizou-se a sequência MPRAGE, extraíndo-se os gânglios da base com o *Freesurfer* e a SB com o algoritmo proposto para segmentação de tecidos do encéfalo. Este último também foi o utilizado para o cálculo das estatísticas da SC, SB e LCR.

A performance do algoritmo não pôde ser validada objetivamente, dado não existirem marcações manuais para os EVR.

8.5.1. Simulação de Espaços de Virchow-Robin

Tentou-se simular EVR para se observar o comportamento do algoritmo numa situação conhecida. Para tal, recorreu-se a uma das imagens da base de dados MRBrainS e utilizou-se a marcação manual para se medir as médias e desvios padrão do LCR, SC e SB, tendo-se simulado cada uma delas com uma distribuição normal. As estatísticas assim calculadas não foram as ideais, pois englobavam não só as intensidades das referidas classes, mas também algum ruído que existisse e mesmo erros de segmentação manual. Assim, não se conseguiu simular a distribuição riciana do ruído. Os EVR foram simulados como segmentos de LCR, com 2 mm de largura, a segmentação dos tecidos era ótima, correspondendo cada *voxel* à sua categoria correta, e as médias e desvios padrão utilizados foram $62,7 \pm 39,9$, $161,1 \pm 32,4$ e $224,9 \pm 15,7$ para LCR, SC e SB, respetivamente. Na Figura 8.14 encontram-se as imagens com os resultados obtidos.

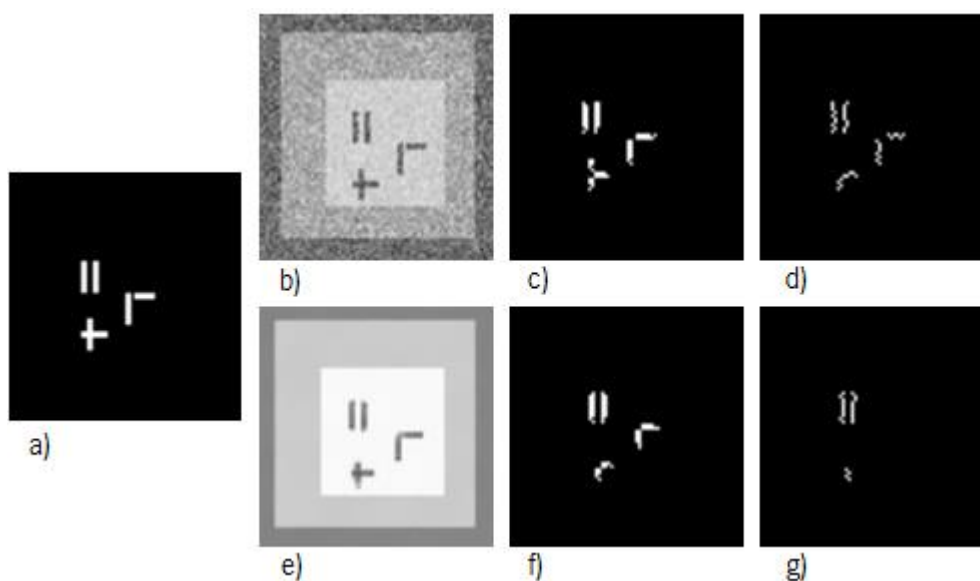


Figura 8.14 - Simulação de Espaços de Virchow-Robin. a) *ground truth* dos EVR simulados; b) simulação; c) candidatos resultantes da combinação dos filtros; d) deteção final dos EVR de b); e) remoção do ruído de b); f) extração de candidatos de e); g) deteção de EVR da imagem e).

Observando a Figura 8.14 constata-se que os filtros falham no realce dos EVR que se cruzam da forma simulada. Isso deve-se ao facto de no ponto de cruzamento o mínimo da vizinhança ser inferior ao do *voxel* central, o que se traduz no resultado negativo da combinação dos filtros, que depois faz com que a procura do comprimento do EVR termine precocemente.

Os candidatos também apresentam zonas onde os EVR apenas medem 1 mm de largura, o que significa que a variação provocada pelo desvio padrão fez com que aí um dos *voxels* representasse melhor as características dos EVR.

Em ambos os casos, o RJMCMC colocou no resultado final os EVR paralelos, dado ser a configuração que sai beneficiada.

Já o resultado para a imagem com remoção de ruído (Figura 8.14(g)) não contém os EVR com configuração de repulsão, indicando que essa configuração era desvantajosa. No entanto, na imagem onde não se removeu ruído essas estruturas estão presentes, e nota-se que o início de ambos está mais distanciado, o que pode fazer com que não se insiram na relação de vizinhança das interações de repulsão. Isto pode ser resultado do movimento *return*, que tenha alterado o sentido dos EVR, e sucessivos *end point move* aceites.

Por seu lado, alguns segmentos que estariam cruzados foram aceites porque o resultado dos filtros fez com que já não existisse interação *hard core*.

No caso em que não se removeu o ruído foram extraídos 74 candidatos a EVR, sendo aceites 21, noutros *slices* que não o do exemplo. No entanto, quando se removeu o ruído o número de candidatos desceu para 32, tendo sido selecionados os 3 visíveis na Figura 8.14(g). Assim sendo, o ruído é responsável por uma elevada taxa de FP, tendo-se considerado positivo o pré-processamento que o corrige, apesar de ter esbatido os *voxels* finais e iniciais dos EVR simulados.

8.5.2. Detecção em imagens reais

Testou-se o algoritmo de deteção automática de EVR nas sequências MPRAGE dos nove pacientes.

Na secção anterior verificou-se que a variabilidade das intensidades, por si só, faz com que existam candidatos em excesso em zonas que por acaso se enquadram nas características de EVR, mesmo com o cálculo das estatísticas para segmentações perfeitas. Assim, como o algoritmo de segmentação de SC, SB e LCR não é ótimo, optou-se por, no cálculo das estatísticas, remover uma certa percentagem dos pontos com intensidades mais baixas e mais elevadas, de forma a tentar compensar os erros de segmentação.

Utilizaram-se 50.000 iterações do algoritmo RJMCMC, e avaliou-se a remoção dos 10% e 20% dos pontos mais e menos intensos de cada classe para o cálculo das estatísticas.

8.5.2.1. Detecção de Espaços de Virchow-Robin nos gânglios da base

Na Figura 8.15 observa-se a energia do modelo para cada percentagem de remoção dos pontos. Verifica-se que 50.000 iterações são suficientes para a convergência do algoritmo em todos os sujeitos. A energia aumenta acentuadamente na primeira metade das iterações, indicando que o modelo atinge o resultado próximo do final nesse espaço, restando as próximas

iterrações para fazer ajustes; a exceção foi o sujeito 4 para a remoção de 20% dos pontos mais e menos intensos, cuja energia após 25.000 iterrações aumentou bruscamente, indicando que existia uma configuração com maior probabilidade, e que talvez a aceitação de um movimento tenha despoletado a aceitação de outros.

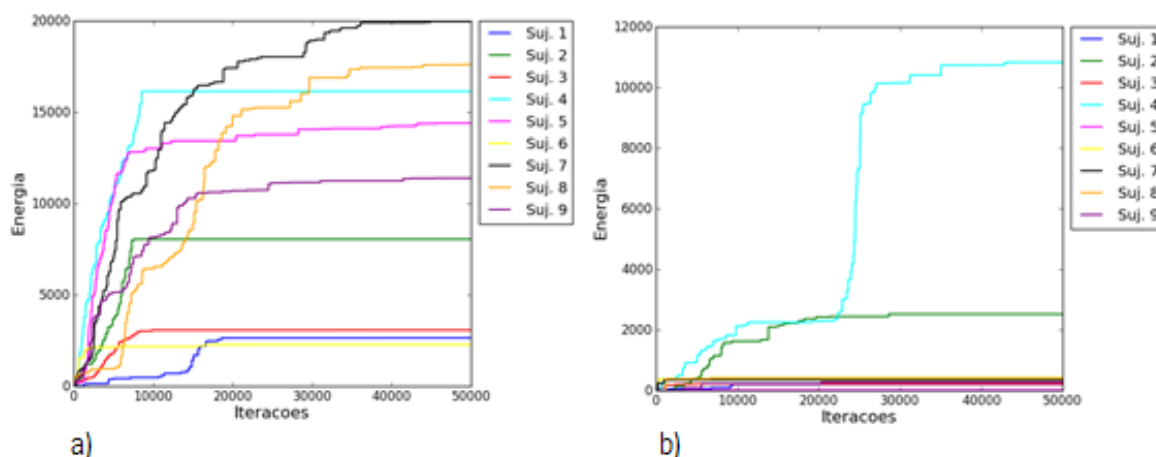


Figura 8.15 - Energia do modelo *Marked Point Process* para Espaços de Virchow-Robin nos gânglios da base. Avaliou-se a remoção dos pontos mais e menos intensos de cada classe para o cálculo das estatísticas, testando a exclusão de a) 10% e b) 20%.

Avaliou-se o número de candidatas a EVR resultantes da aplicação dos filtros, e aqueles que foram aceites pelo RJMCMC, podendo-se consultar na Tabela 8.10. Verificou-se que nos gânglios da base são detetados muitos candidatos, pois esta região é constituída por SC, o que pode fazer com que a variabilidade natural da intensidade dos tecidos faça com que algumas regiões obtenham respostas positivas nos filtros. No entanto, o RJMCMC é capaz de eliminar uma grande quantidade de candidatas, escolhendo os que melhor se enquadram no modelo de EVR.

A remoção de 20% dos pontos mais e menos intensos resulta em menos candidatas, mas tem maior impacto ainda sobre o número de segmentos aceites pelo RJMCMC. Isto deve-se à diminuição do desvio padrão de cada classe, o que aumenta o declive da resposta dos filtros, e redução da gama de intensidades de respostas positivas.

O sujeito 4 é aquele para o qual se detetam mais EVR, particularmente quando se removem 20% dos pontos nas estatísticas, explicando a maior energia face aos outros.

Tabela 8.10 - Número de candidatos a EVR e segmentos aceites nos gânglios da base, para cada percentagem de remoção de pontos.

% removida	EVR	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Suj. 6	Suj. 7	Suj. 8	Suj. 9
10	Candidatos	757	1130	410	1004	892	146	853	752	916
	Aceites	23	38	28	90	64	20	80	70	69
20	Candidatos	533	620	185	574	473	41	601	358	646
	Aceites	3	26	5	49	2	3	4	2	2

Fez-se a contagem do número de segmentos aceites por cada comprimento, em milímetros arredondados às unidades, tal como se pode observar na Figura 8.16. Verificou-se que a maior parte dos segmentos aceites medem 3 mm, correspondendo ao mínimo da dimensão permitida. Após o aumento da quantidade de pontos removidos, a maioria dos segmentos aceites ainda mede 3mm, no entanto existe uma proporção maior das outras medidas, indiciando que há muitos FPs com o comprimento mínimo, correspondendo a pequenas quantidades de *voxels* que, aleatoriamente, se enquadram em alguns pressupostos para os EVR.

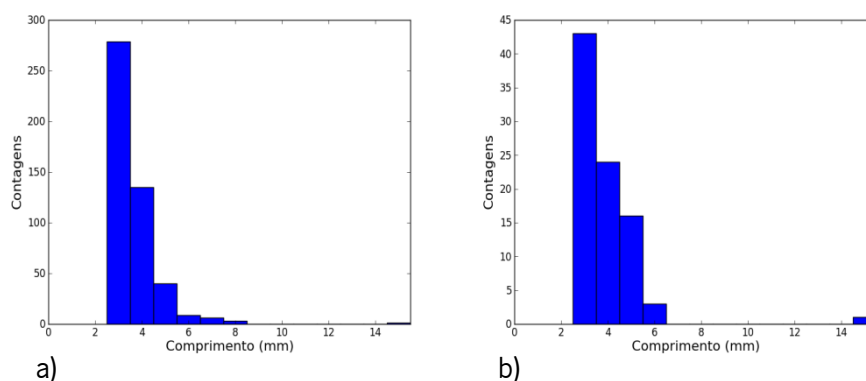


Figura 8.16 - Quantidade de Espaços de Virchow-Robin aceites por comprimento, nos gânglios da base. Em a) as estatísticas foram calculadas após remoção de 10% dos pontos mais e menos intensos, enquanto em b) a quantidade foi de 20%.

Na Figura 8.17 observam-se os EVR detetados nos gânglios da base de um dos sujeitos, e a região de interesse que realmente foi usada no processamento.

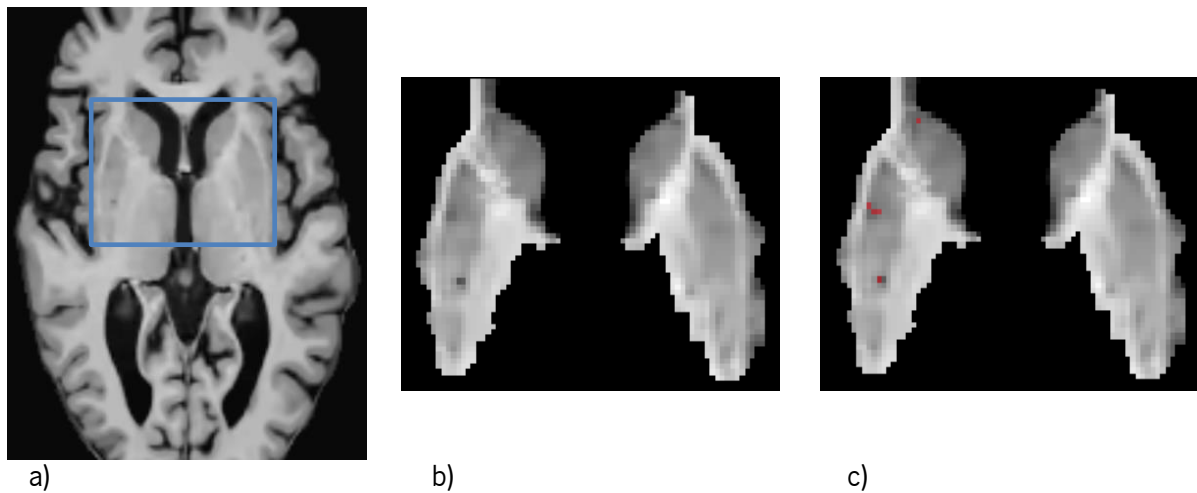


Figura 8.17 - Espaços de Virchow-Robin nos gânglios da base. a) plano axial do encéfalo com a região de interesse envolvida pelo retângulo azul; b) gânglios da base; c) gânglios da base com os EVR detetados marcados.

Na Figura 8.18 podem-se observar alguns exemplos de EVR aceites pelo RJMCMC. De facto, são detetados segmentos em regiões hipointensas, que poderão ter algumas características das estruturas de interesse.

No entanto, na Figura 8.18(c) foi aceite um candidato num local que aparentemente não apresenta nenhuma característica anormal, mas existem *voxels* que são um pouco menos intensos do que a média da SC, e a vizinhança considerada pelos filtros é um pouco mais clara, resultando em respostas reduzidas dos filtros, mas positivas. Na Figura 8.18(f) deu-se uma situação parecida, mas aí as intensidades dos *voxels* são, mesmo visualmente, menos intensas do que a restante SC, ou seja, o aumento da percentagem de pontos removidos melhora o algoritmo em relação ao número de FPs, mas não é suficiente para colmatar o problema. Assim sendo, o algoritmo demonstrou potencial para ser aplicado somente nos gânglios da base, mas deve-se estudar modificações nos filtros que os tornem mais específicos para a SC.

8.5.2.2. Detecção de Espaços de Virchow-Robin na substância branca

Aplicou-se o algoritmo de detecção automática de EVR na SB cerebral, encontrando-se na Figura 8.19 a energia do modelo ao longo das iterações, e na Tabela 8.11 o número de candidatos a EVR e aqueles que foram aceites.

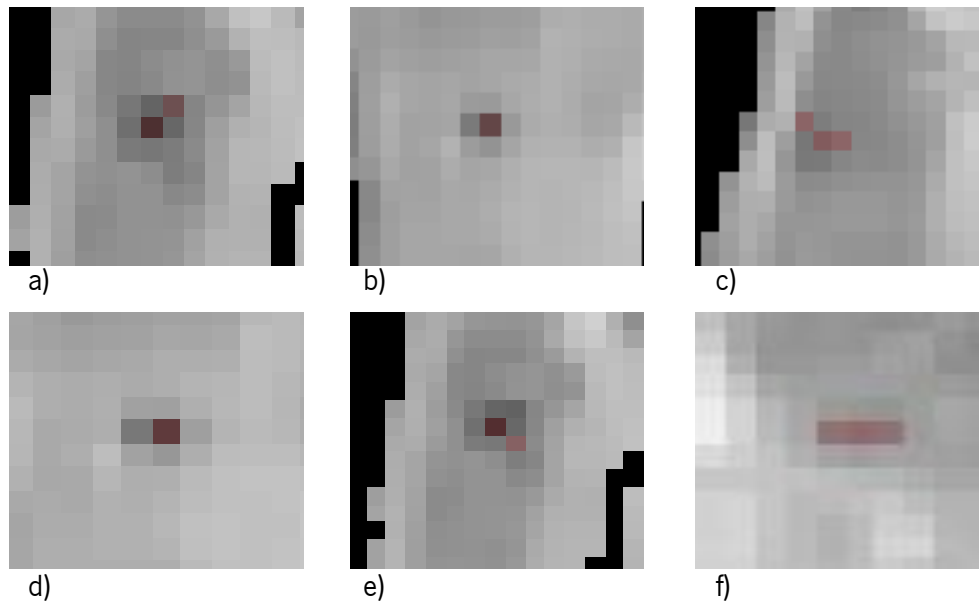


Figura 8.18 - Espaços de Virchow-Robin detetados nos gânglios da base. Na primeira linha encontram-se exemplos de EVR detetados quando se exclui 10% dos *voxels* com maior e menor intensidade no cálculo das estatísticas, enquanto na segunda os exemplos referem-se ao caso em que a quantidade de remoção foi de 20%.

Neste caso, também 50.000 iterações foi suficiente para a convergência do RJMCMC, verificando-se que, com exceção do sujeito 4, esse estado foi atingido antes das 10.000 iterações. Este facto deve-se ao menor número de candidatos a EVR, relativamente aos gânglios da base, o que reduz as possibilidades de procura do algoritmo e possíveis “saltos” entre dimensões.

Analisando a Tabela 8.11 verifica-se que o número de candidatos a EVR e segmentos aceites é significativamente inferior ao dos gânglios da base. Este facto deve-se à SB contrastar mais com os EVR, e a sua intensidade estar mais distante dessas estruturas, o que faz com que não exista tanto tecido normal que possa estar em zonas de resposta positiva dos filtros.

A remoção de 20% dos pontos mais e menos intensos durante o cálculo das estatísticas resultou em menos candidatos e menos EVR aceites, devido à diminuição da gama de intensidades para as quais os filtros obtinham resposta positiva.

A contagem do número de candidatos aceites pelo RJMCMC por comprimento, em milímetros arredondados às unidades, pode ser consultada na Figura 8.20. Tal como nos gânglios da base, também na SB a maioria dos segmentos detetados media 3 mm, e a remoção de uma maior quantidade de pontos para o cálculo das estatísticas teve maior efeito sobre esses.

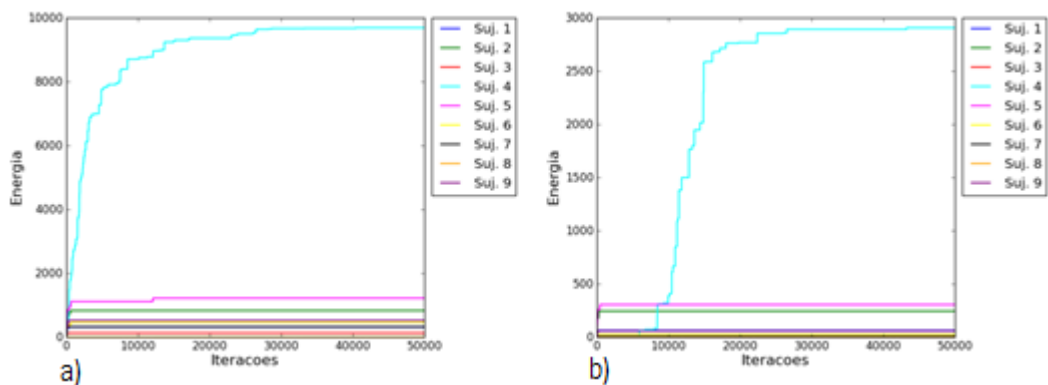
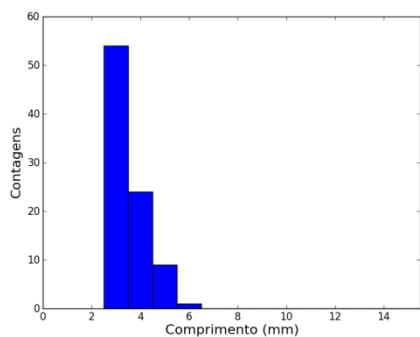


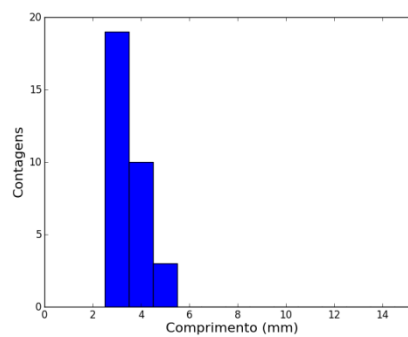
Figura 8.19 - Energia do modelo *Marked Point Process* para Espaços de Virchow-Robin na substância branca cerebral. Avaliou-se a remoção de 10% dos pontos mais e menos intensos de cada classe para o cálculo das estatísticas, testando a exclusão de a) 10% e b) 20%.

Tabela 8.11 - Número de candidatos a Espaço de Virchow-Robin e segmentos aceites na substância branca cerebral, para cada percentagem de remoção de pontos.

% removida	EVR	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Suj. 6	Suj. 7	Suj. 8	Suj. 9
10	Candidatos	10	43	5	632	32	28	18	2	28
	Aceites	4	7	2	59	10	6	3	0	5
20	Candidatos	6	29	0	388	18	11	12	2	10
	Aceites	1	5	0	26	3	1	1	0	2



a)



b)

Figura 8.20 - Quantidade de Espaços de Virchow-Robin aceites por comprimento, na substância branca cerebral. Em a) as estatísticas foram calculadas após remoção de 10% dos pontos mais e menos intensos, enquanto em b) a quantidade foi de 20%.

Na Figura 8.21 observam-se os EVR detetados na SB de um dos sujeitos, e uma aproximação à região onde foram detetadas duas dessas estruturas, na região de interesse utilizada no processamento.

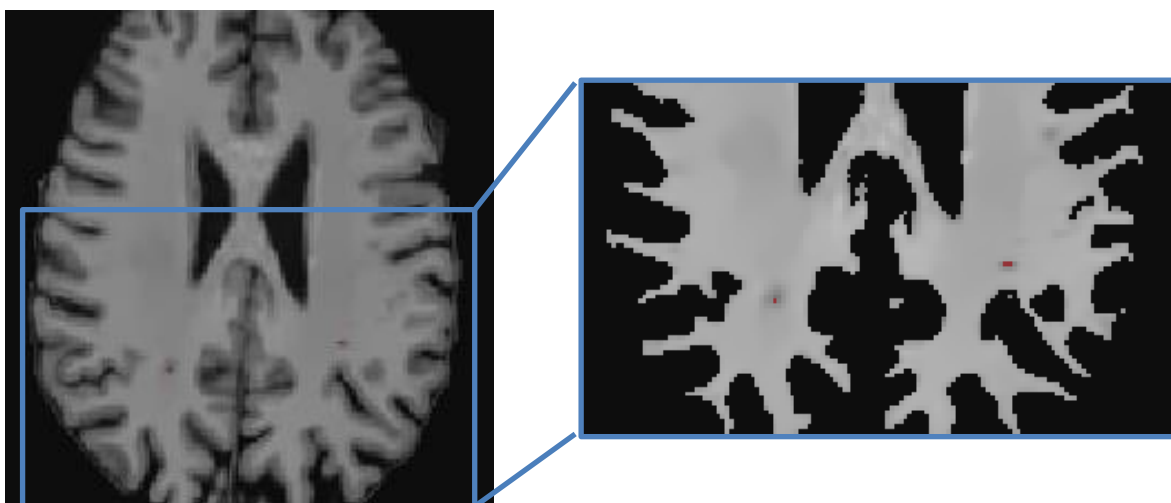


Figura 8.21 - Espaços de Virchow-Robin na substância branca.

Observando os exemplos de EVR aceites pelo RJMCMC da Figura 8.22, constata-se que é capaz de detetar objetos em regiões com características dessas estruturas, nomeadamente por conterem *voxels* com menores intensidades, próximas do LCR.

Na Figura 8.22(c) e na Figura 8.22(f) encontram-se exemplos de EVR aceites para o sujeito 4. Este foi aquele com maior número de candidatos e estruturas aceites, sendo a maioria FPs localizados em grandes áreas de tecido que pode estar lesionado, uma vez que demonstram ter intensidades mais baixas. Assim, pode-se pensar que a remoção de mais pontos no cálculo das estatísticas teve maior impacto nestas regiões, dado que o número de candidatos aceites reduziu para menos de metade.

Na globalidade, o algoritmo parece detetar melhor os EVR na SB do que nos gânglios da base, não sendo necessário excluir uma grande quantidade de pontos para se obter poucos candidatos e segmentos FPs.

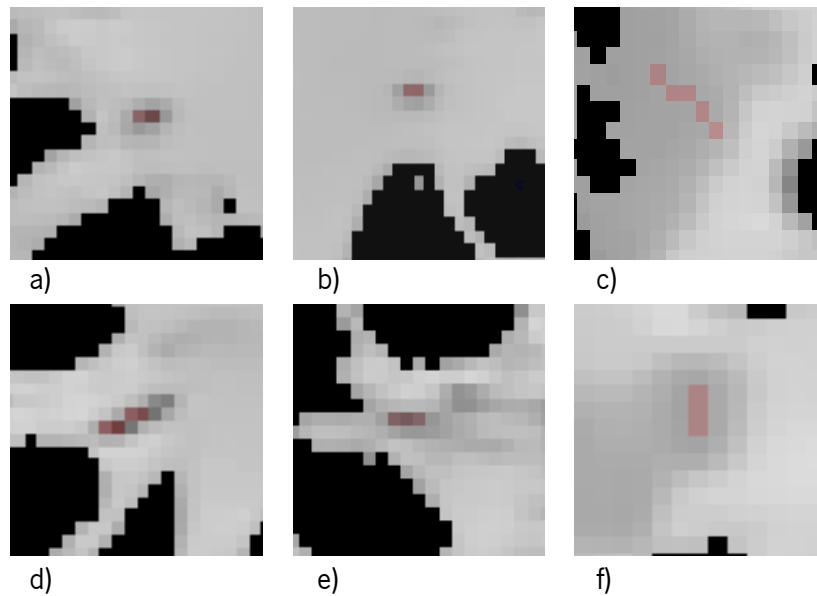


Figura 8.22 - Espaços de Virchow-Robin detetados na substância branca cerebral. Na primeira linha encontram-se exemplos de EVR detetados quando se exclui 10% dos *voxels* com maior e menor intensidade no cálculo das estatísticas, enquanto na segunda os exemplos referem-se ao caso em que a quantidade de remoção foi de 20%.

8.6. Sumário

Neste capítulo apresentou-se os resultados obtidos ao longo da presente dissertação.

Assim, para a segmentação de SC, SB e LCR, avaliaram-se as características utilizadas, verificando-se que as mais relevantes foram as que incluíam informação da vizinhança. Além disso, concluiu-se que o coeficiente de Gini é equivalente, neste contexto, ao ganho de informação para avaliar a divisão dos nós das RFs, que a utilização de 100.000 extraídos aleatoriamente de cada sujeito é suficiente para treinar o classificador, e que o aumento do número de DTs é positivo, mas no caso do número de nós de profundidade máxima pode ser negativo, diminuindo os resultados a partir de determinado momento devido ao *overfitting*. Assim sendo, o melhor resultado, em Dice, deu-se com uma floresta com 60 DTs e 25 nós de profundidade máxima permitida, obtendo-se $0,719 \pm 0,019$, $0,847 \pm 0,014$ e $0,885 \pm 0,023$ para LCR, SC e SB, respetivamente, no conjunto de treino utilizando *leave-one-out*, e $0,683 \pm 0,039$, $0,834 \pm 0,026$ e $0,876 \pm 0,017$ no conjunto de teste.

Uma vez que apenas se iria utilizar a sequência T1 para a deteção de EVR, avaliou-se o impacto da utilização apenas desse tipo de imagens sobre a segmentação de tecidos, obtendo-se os resultados, em Dice, para o conjunto de treino com *leave-one-out* de $0,713 \pm 0,019$, $0,837 \pm 0,017$ e $0,878 \pm 0,026$ para LCR, SC e SB, respetivamente. Assim se verificou que T1 é a mais importante para a distinção entre as classes, mas a FLAIR é muito relevante se existirem

lesões da SB, que neste caso foram mal classificadas como SC, quando anteriormente eram corretamente atribuídas à SB. Utilizando-se este treino na base de dados MRBrainS, aplicou-se o algoritmo às sequências MPRAGE disponíveis, e verificou-se que, devido à normalização, eram corretamente segmentadas.

Na detecção dos EVR não se pôde fazer uma validação objetiva, dado não existirem marcações manuais. No entanto, observou-se que o algoritmo é mais eficaz quando aplicado na SB, podendo ser aplicado nos gânglios da base mas o número de FP aumenta, pelo que se deve investigar alterações nos filtros que os melhorem para regiões de SC. O algoritmo RJMCMC atinge o estado de convergência, tendo-se verificado que 50.000 iterações são suficientes para estas regiões de interesse.

Capítulo 9 Conclusão e trabalho futuro

Ao longo deste capítulo irá sintetizar-se o trabalho desenvolvido e expor-se as conclusões mais relevantes. Em seguida propor-se-á alguns caminhos para melhorar o sistema desenvolvido ou testar outras abordagens, em trabalho futuro.

9.1. Conclusões

Os EVR rodeiam as artérias perfurantes do encéfalo, encontrando-se preenchidos por LCR e líquido intersticial, e apresentando-se com reduzidas dimensões. A sua observação em RM tem sido apontada como um biomarcador de risco de algumas condições de saúde, como o risco de AVC ou demência vascular, tendo sido utilizadas escalas semi-quantitativas nesses estudos. Essas correlações ganham mais significado quando se restringe a observação a regiões de interesse, em particular os gânglios da base e a SB cerebral [4], [30].

Assim sendo, o principal objetivo da dissertação era a implementação de um sistema que fizesse a deteção e contagem de EVR nessas regiões de interesse. Para tal, foi necessário implementar uma *pipeline* que aplicasse os necessários pré-processamentos. Desta forma, utilizou-se implementações já existentes para esta fase, nomeadamente para corrigir a falta de homogeneidade [76], remoção do crânio [79] e ruído [86], e normalização das intensidades [80] para utilização no método supervisionado. Concluiu-se que os procedimentos adotados são adequados, dando-se destaque à remoção de ruído, pois este é significativo nas sequências MPRAGE e resulta na extração de candidatos a EVR em excesso, e conseqüentemente na aceitação de FPs.

Para a extração dos gânglios da base aplicou-se o método disponibilizado pelo *Freesurfer*, aplicando-se depois um pós-processamento que tornou a sua segmentação apropriada para a aplicação em causa.

Já a SB cerebral foi extraída com um algoritmo original, usando RFs, que demonstrou ser adequado para a segmentação de SC e SB, sendo robusto à presença de lesões da SB se a sequência FLAIR se encontrar disponível. A sua aplicabilidade na segmentação de sequências

MPRAGE também foi demonstrada, verificando-se a importância do procedimento de normalização das intensidades. Relativamente às características utilizadas, concluiu-se que no contexto de segmentação de tecidos, a informação da vizinhança é muito importante.

Depois da implementação do algoritmo de detecção de EVR proposto em [24], concluiu-se que o mesmo é capaz de selecionar entre os candidatos alguns que se adequem aos pressupostos de EVR, parecendo conseguir melhores performances para a SB cerebral. Concluiu-se ainda que é sensível ao resultado da segmentação que lhe permite calcular estatísticas dos tecidos, e à presença de lesões na SB, mas ainda assim o algoritmo RJMCMC converge para uma configuração, verificando-se que 50.000 iterações são suficientes para ambas as regiões de interesse. Porém, não pôde ser realizada uma validação objetiva, dado não existirem marcações manuais.

Por fim, desenvolveu-se uma aplicação que permite visualizar os EVR detetados, e analisar a forma 3D da progressão destes através do tecido. Além disso, através de reconstrução 3D pode-se observar a região de interesse extraída, verificando-se se foi correta. Também se visualiza os EVR detetados em sobreposição à região de interesse, e a reconstrução destes no espaço, o que permite analisar a sua distribuição espacial.

9.2. Trabalho futuro

Dado o tempo limitado para o desenvolvimento do sistema proposto, não se pôde testar mais alternativas, ou aperfeiçoar alguns procedimentos, pelo que se pode considerar trabalho futuro.

Assim, na fase de pré-processamento deve-se melhorar a remoção do crânio, quer seja através da pesquisa de outras alternativas, quer seja propondo uma técnica nova. Desta forma, acredita-se que a segmentação do encéfalo melhorará, em particular o LCR.

Relativamente à extração dos gânglios da base, poderia investigar-se a possibilidade de utilizar RFs para os segmentar. Neste sentido, propor-se-ia, ou pesquisar-se-ia, novas características que melhorem o treino com 6 classes e evitem que essas regiões sejam classificadas como córtex cerebral; ou investigar-se-ia a possibilidade de utilizar o atlas como forma de reduzir a região de interesse a segmentar, e aplicar-se aí o classificador.

Na segmentação de SC, SB e LCR poder-se-ia começar por aplicar um algoritmo de seleção de características, que dispensasse as menos relevantes, de forma sistematizada, apesar de a própria RF ser capaz de lidar com essas, tal como se verificou quando se utilizou

apenas T1. Para além de uma remoção do crânio superior como forma de melhoria do LCR, deveria procurar-se características que forneçam informação das regiões com mais detalhes como os sulcos e giros, o que se traduziria numa segmentação superior da SC e SB. Já no que se refere à utilização sobre a sequência MPRAGE, poder-se-ia avaliar o efeito sobre as lesões da SB de registar e interpolar a sequência FLAIR com distância entre *slices* de 6 mm, ou tentar obter aquisições com essa distância menor.

Finalmente, relativamente ao algoritmo de deteção de EVR, deveria começar-se por avaliar a sua performance comparativamente a marcações manuais. No entanto, entre os possíveis trabalhos futuros poder-se-ia apontar a projeção de filtros específicos para os gânglios da base, a melhoria do cálculo das estatísticas de tecidos que passa pelo aumento da performance do método utilizado, ou a avaliação da utilização de um algoritmo supervisionado de seleção dos candidatos, ou pré-seleção antes do MPP e RJMCMC.

Bibliografia

- [1] K. H. Elaine N. Marieb, *Human Anatomy & Physiology*, 7th Editio. Benjamin-Cummings Publishing Company, 2006.
- [2] R. M. Kwee and T. C. Kwee, "Virchow-Robin spaces at MR imaging.," *Radiographics*, vol. 27, no. 4, pp. 1071–86, 2007.
- [3] R. P. W. Rouhl, R. J. van Oostenbrugge, I. L. H. Knottnerus, J. E. a Staals, and J. Lodder, "Virchow-Robin spaces relate to cerebral small vessel disease severity.," *J. Neurol.*, vol. 255, no. 5, pp. 692–6, May 2008.
- [4] T. F. Patankar, D. Mitra, A. Varma, J. Snowden, D. Neary, and A. Jackson, "Dilatation of the Virchow-Robin space is a sensitive indicator of cerebral microvascular disease: study in elderly patients with dementia.," *AJNR. Am. J. Neuroradiol.*, vol. 26, no. 6, pp. 1512–20, 2005.
- [5] S. Sacco, C. Marini, R. Totaro, T. Russo, D. Cerone, and A. Carolei, "A population-based study of the incidence and prognosis of lacunar stroke," *Neurology*, vol. 66, no. 9, pp. 1335–1338, 2006.
- [6] Y. Bejot, A. Catteau, M. Caillier, O. Rouaud, J. Durier, C. Marie, A. DI Carlo, G. Osseby, T. Moreau, and M. Giroud, "Trends in Incidence, Risk Factors, and Survival in Symptomatic Lacunar Stroke in Dijon, France, From 1989 to 2006 A Population-Based Study," *Stroke*, vol. 39, no. 7, pp. 1945–1951, 2008.
- [7] L. Battistin and A. Cagnin, "Vascular cognitive disorder. A biological and clinical overview," *Neurochem. Res.*, vol. 35, pp. 1933–1938, 2010.
- [8] M. D. C. V. Hernández, R. J. Piper, X. Wang, I. J. Deary, and J. M. Wardlaw, "Towards the automatic computational assessment of enlarged perivascular spaces on brain magnetic resonance images: A systematic review.," *J. Magn. Reson. Imaging*, vol. 000, Feb. 2013.
- [9] Z. Yi, A. Criminisi, J. Shotton, and A. Blake, "Discriminative, semantic segmentation of brain tissue in MR images.," *Int. Conf. Med. Image Comput. Comput. Interv.*, vol. 12, no. Pt 2, pp. 558–65, Jan. 2009.
- [10] D. Zikic, B. Glocker, E. Konukoglu, and J. Shotton, "Context-sensitive Classification Forests for Segmentation of Brain Tumor Tissues," *Int. Conf. Med. Image Comput. Comput. Interv.*, 2012.
- [11] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, and T. P. Roberts, "MR pulse sequences: what every radiologist wants to know but is afraid to ask.," *Radiographics*, vol. 26, no. 2, pp. 513–37, 2006.

- [12] P. Rinck, *Magnetic Resonance in Medicine. The Basic Textbook of the European Magnetic Resonance Forum*, 6th ed. 2012.
- [13] P. Brodal, *The Central Nervous System: Structure and Function*. New York, 2004.
- [14] S. I. Fox, *Human Physiology*, 12th ed. New York City: McGraw-Hill Higher Education, 2010.
- [15] R. A. B. H. Royden Jones Jr., Jayashri Srinivasan, Gregory J. Allam, *Netter's Neurology*, 2nd Editio. Elsevier Inc., 2012.
- [16] A. Mechelli and C. Price, "Voxel-based morphometry of the human brain: methods and applications," *Curr. Med. Imaging Rev.*, vol. 1, no. 2, pp. 105–113, 2005.
- [17] S. Sigurdsson, T. Aspelund, and L. Forsberg, "Brain tissue volumes in the general population of the elderly: the AGES-Reykjavik study," *Neuroimage*, vol. 59, no. 4, pp. 3862–3870, 2012.
- [18] M. Lehmann, S. Crutch, and G. Ridgway, "Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease," *Neurobiol. Aging*, vol. 32, no. 8, pp. 1466–1476, 2011.
- [19] A. R. Stoffers, D., Sheldon, S., Kuperman, J.M., Goldstein, J., Corey-Bloom, J., Aron, "Contrasting gray and white matter changes in preclinical Huntington disease," *Neurology*, vol. 74, no. 15, pp. 1208–1216, 2010.
- [20] R. Honea, T. J. Crow, D. Passingham, and C. E. Mackay, "Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies," *Am. J. Psychiatry*, vol. 162, no. 12, pp. 2233–2245, 2005.
- [21] J. Wuerfel, M. Haertle, H. Waiczies, E. Tysiak, I. Bechmann, K. D. Wernecke, F. Zipp, and F. Paul, "Perivascular spaces—MRI marker of inflammatory activity in the brain?," *Brain*, vol. 131, no. Pt 9, pp. 2332–40, Sep. 2008.
- [22] S. Groeschel, W. K. Chong, R. Surtees, and F. Hanefeld, "Virchow-Robin spaces on magnetic resonance images: normative data, their dilatation, and a review of the literature.," *Neuroradiology*, vol. 48, no. 10, pp. 745–54, Oct. 2006.
- [23] I. Bechmann, I. Galea, and V. H. Perry, "What is the blood-brain barrier (not)?," *Trends Immunol.*, vol. 28, no. 1, pp. 5–11, Jan. 2007.
- [24] X. Descombes, F. Kruggel, G. Wollny, and H. J. Gertz, "An object-based approach for detecting small brain lesions: application to Virchow-Robin spaces.," *IEEE Trans. Med. Imaging*, vol. 23, no. 2, pp. 246–55, Feb. 2004.
- [25] K. Salzman and A. Osborn, "Giant tumefactive perivascular spaces," *Am. J. Neuroradiol.*, vol. 26, no. February, pp. 298–305, 2005.

- [26] O. Fayeye, B. L. Pettorini, K. Foster, and D. Rodrigues, "Mesencephalic enlarged Virchow-Robin spaces in a 6-year-old boy: a case-based update.," *Childs. Nerv. Syst.*, vol. 26, no. 9, pp. 1155–60, Sep. 2010.
- [27] N. Benhaïem-Sigaux, F. Gray, R. Gherardi, a. M. Roucayrol, and J. Poirier, "Expanding cerebellar lacunes due to dilatation of the perivascular space associated with Binswanger's subcortical arteriosclerotic encephalopathy," *Stroke*, vol. 18, no. 6, pp. 1087–1092, Nov. 1987.
- [28] P. Homeyer, P. Cornu, L. Lacomblez, and J. Chiras, "SHORT REPORT A special form of cerebral lacunae : expanding lacunae," *J. Neurol. Neurosurg. Psychiatry*, vol. 61, pp. 200–202, 1996.
- [29] H. Pollock, M. Hutchings, R. O. Weller, and E. T. Zhang, "Perivascular spaces in the basal ganglia of the human brain: their relationship to lacunes.," *J. Anat.*, vol. 191 (Pt 3, pp. 337–46, Oct. 1997.
- [30] F. N. Doubal, A. M. J. MacLulich, K. J. Ferguson, M. S. Dennis, and J. M. Wardlaw, "Enlarged perivascular spaces on MRI are a feature of cerebral small vessel disease.," *Stroke*, vol. 41, no. 3, pp. 450–4, Mar. 2010.
- [31] V. Hachinski, C. Iadecola, R. C. Petersen, M. M. Breteler, D. L. Nyenhuis, S. E. Black, W. J. Powers, C. DeCarli, J. G. Merino, R. N. Kalara, H. V Vinters, D. M. Holtzman, G. a Rosenberg, A. Wallin, M. Dichgans, J. R. Marler, and G. G. Leblanc, "National Institute of Neurological Disorders and Stroke-Canadian Stroke Network vascular cognitive impairment harmonization standards.," *Stroke*, vol. 37, no. 9, pp. 2220–41, Sep. 2006.
- [32] J. M. Wardlaw, E. E. Smith, G. Biessels, C. Cordonnier, F. Fazekas, R. Frayne, R. I. Lindley, J. T. O'Brien, F. Barkhof, O. R. Benavente, S. E. Black, C. Brayne, M. Breteler, H. Chabriat, C. DeCarli, F.-E. de Leeuw, F. Doubal, M. Duering, N. C. Fox, S. Greenberg, V. Hachinski, I. Kilimann, V. Mok, R. van Oostenbrugge, L. Pantoni, O. Speck, B. C. M. Stephan, S. Teipel, A. Viswanathan, D. Werring, C. Chen, C. Smith, M. van Buchem, B. Norrving, P. B. Gorelick, and M. Dichgans, "Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration," *Lancet Neurol.*, vol. 12, no. 8, pp. 822–838, 2013.
- [33] C. J. Song, J. H. Kim, E. L. Kier, and R. a Bronen, "MR imaging and histologic features of subinsular bright spots on T2-weighted MR images: Virchow-Robin spaces of the extreme capsule and insular cortex.," *Radiology*, vol. 214, no. 3, pp. 671–7, Mar. 2000.
- [34] L. Pantoni, "Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges.," *Lancet Neurol.*, vol. 9, no. 7, pp. 689–701, Jul. 2010.
- [35] Y.-C. Zhu, C. Tzourio, A. Soumaré, B. Mazoyer, C. Dufouil, and H. Chabriat, "Severity of dilated Virchow-Robin spaces is associated with age, blood pressure, and MRI markers of small vessel disease: a population-based study.," *Stroke*, vol. 41, no. 11, pp. 2483–90, Dec. 2010.

- [36] a M. J. MacLulich, J. M. Wardlaw, K. J. Ferguson, J. M. Starr, J. R. Seckl, and I. J. Deary, "Enlarged perivascular spaces are associated with cognitive function in healthy elderly men.," *J. Neurol. Neurosurg. Psychiatry*, vol. 75, no. 11, pp. 1519–23, Nov. 2004.
- [37] T. F. Patankar, R. Baldwin, D. Mitra, S. Jeffries, C. Sutcliffe, A. Burns, and A. Jackson, "Virchow-Robin space dilatation may predict resistance to antidepressant monotherapy in elderly patients with depression.," *J. Affect. Disord.*, vol. 97, no. 1–3, pp. 265–70, Jan. 2007.
- [38] D. L. Pham and J. L. Prince, "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recognit. Lett.*, vol. 20, no. 1, pp. 57–68, Jan. 1999.
- [39] A. Liew and H. Yan, "An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Trans. Med. Imaging*, vol. 22, no. 9, pp. 1063–1075, 2003.
- [40] H. Wang and B. Fei, "A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme," *Med. Image Anal.*, vol. 13, pp. 193–202, 2009.
- [41] M. Cardoso, M. Clarkson, and G. Ridgway, "LoAd: A locally adaptive cortical segmentation algorithm," *Neuroimage*, vol. 56, pp. 1386–1397, 2011.
- [42] W. M. Wells, W. L. Grimson, R. Kikinis, and F. a Jolesz, "Adaptive segmentation of MRI data.," *IEEE Trans. Med. Imaging*, vol. 15, no. 4, pp. 429–42, Jan. 1996.
- [43] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain.," *IEEE Trans. Med. Imaging*, vol. 18, no. 10, pp. 897–908, Oct. 1999.
- [44] 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. Imaging*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [45] J. Ashburner and K. J. Friston, "Unified segmentation.," *Neuroimage*, vol. 26, no. 3, pp. 839–51, Jul. 2005.
- [46] H. Greenspan, "Constrained Gaussian mixture model framework for automatic segmentation of MR brain images," *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1233–1245, 2006.
- [47] D. Pham and J. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Med. Imaging*, vol. 18, no. 9, pp. 737–752, 1999.
- [48] Z. Song, N. Tustison, B. Avants, and J. Gee, "Integrated graph cuts for brain MRI segmentation," *Int. Conf. Med. Image Comput. Comput. Interv.*, vol. 4191, pp. 831–838, 2006.

- [49] A. Mayer and H. Greenspan, "An adaptive mean-shift framework for MRI brain segmentation," *IEEE Trans. Med. Imaging*, vol. 28, no. 8, pp. 1238–1250, 2009.
- [50] J. Jiménez-Alaniz, "Data-driven brain MRI segmentation supported on edge confidence and a priori tissue information," *IEEE Trans. Med. Imaging*, vol. 25, no. 1, pp. 74–83, 2006.
- [51] P. Anbeek and K. Vincken, "Probabilistic segmentation of brain tissue in MR imaging," *Neuroimage*, vol. 27, pp. 795–804, 2005.
- [52] A. Akselrod-Ballin and M. Galun, "An integrated segmentation and classification approach applied to multiple sclerosis analysis," *Comput. Vis. Pattern Recognition, 2006 IEEE Comput. Soc. Conf.*, vol. 1, pp. 1122–1129, 2006.
- [53] A. Akselrod-Ballin, M. Galun, and M. Gomori, "Atlas guided identification of brain structures by combining 3D segmentation and SVM classification," *Lect. Notes Comput. Sci. Med. Image Comput. Comput. Interv. – MICCAI 2006*, vol. 4191, pp. 209–216, 2006.
- [54] Y. Uchiyama, T. Kunieda, T. Asano, H. Kato, T. Hara, M. Kanematsu, T. Iwama, H. Hoshi, Y. Kinoshita, and H. Fujita, "Computer-aided diagnosis scheme for classification of lacunar infarcts and enlarged Virchow-Robin spaces in brain MR images.," *Eng. Med. Biol. Soc. 2008. EMBS 2008. 30th Annu. Int. Conf. IEEE*, vol. 2008, pp. 3908–11, Jan. 2008.
- [55] J. Ramirez, E. Gibson, a Quddus, N. J. Lobaugh, a Feinstein, B. Levine, C. J. M. Scott, N. Levy-Cooperman, F. Q. Gao, and S. E. Black, "Lesion Explorer: a comprehensive segmentation and parcellation package to obtain regional volumetrics for subcortical hyperintensities and intracranial tissue.," *Neuroimage*, vol. 54, no. 2, pp. 963–73, Jan. 2011.
- [56] 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. Imaging*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [57] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester, UK: John Wiley & Sons, Ltd, 2007.
- [58] A. Baddeley, "Spatial point processes and their applications," in in *Stochastic Geometry*, vol. 3, Lectures given at the C.I.M.E. Summer School held in Martina Franca, Italy, 2007.
- [59] R. Stoica, X. Descombes, and J. Zerubia, "A Gibbs Point Process for Road Extraction from Remotely Sensed Images," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 121–136, 2004.
- [60] R. Stoica, X. Descombes, and J. Zérubia, "Road extraction in remote sensed images using stochastic geometry framework," *AIP Conf. Proc.*, vol. 568, no. 531, pp. 531–542, 2001.
- [61] C. Andrieu, N. De Freitas, A. Doucet, and M. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.*, vol. 50, no. 1–2, pp. 5–43, 2003.

- [62] N. Metropolis and S. Ulam, "The monte carlo method," *J. Am. Stat. Assoc.*, vol. 44, no. 247, pp. 335–341, 1949.
- [63] P. J. Green, "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, vol. 82, no. 4, p. 711, Dec. 1995.
- [64] C. Geyer, "Markov chain Monte Carlo lecture notes," *Course notes, Spring Quart.*, 1998.
- [65] K. P. Murphy, *Machine Learning - A Probabilistic Perspective*, 1st ed. The MIT Press, 2012.
- [66] P. Green and D. Hastie, "Reversible jump MCMC," *Genetics*, vol. 155, no. 3, pp. 1391–1403, 2009.
- [67] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science (80-)*, vol. 220, no. 4598, pp. 671–680, 1983.
- [68] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4ª Edição. Londres: Academic Press Inc, 2009.
- [69] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Int. Jt. Conf. Artif. Intell. - IJCAI*, 1995.
- [70] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," 2003.
- [71] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning," *Found. Trends@ Comput. Graph. Vis.*, vol. 7, no. 2–3, pp. 81–227, 2011.
- [72] K. Gorgolewski, C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. S. Ghosh, "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python.," *Front. Neuroinform.*, vol. 5, no. August, p. 13, Jan. 2011.
- [73] U. Vovk, F. Pernus, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI.," *IEEE Trans. Med. Imaging*, vol. 26, no. 3, pp. 405–21, Mar. 2007.
- [74] 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. Imaging*, vol. 17, no. 1, pp. 87–97, Feb. 1998.
- [75] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, pp. 782–790, 2012.
- [76] N. J. Tustison, B. B. Avants, P. a Cook, Y. Zheng, A. Egan, P. a Yushkevich, and J. C. Gee, "N4ITK: improved N3 bias correction.," *IEEE Trans. Med. Imaging*, vol. 29, no. 6, pp. 1310–20, Jun. 2010.

- [77] B. B. Avants, N. Tustison, and G. Song, "Advanced Normalization Tools (ANTs)," *Insight J.*, 2009.
- [78] S. M. Smith, "Fast robust automated brain extraction.," *Hum. Brain Mapp.*, vol. 17, no. 3, pp. 143–55, Nov. 2002.
- [79] F. Ségonne, 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*, vol. 22, no. 3, pp. 1060–75, Jul. 2004.
- [80] L. Nyúl, J. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imaging*, vol. 19, no. 2, pp. 143–150, 2000.
- [81] T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, "Engineering and Algorithm Design for an Image Processing API: A Technical Report on ITK - The Insight Toolkit," *Proc. Med. Meets Virtual Reality, J. Westwood, ed., IOS Press Amsterdam*, pp. 586–592, 2002.
- [82] L. G. Nyúl and J. K. Udupa, "On standardizing the MR image intensity scale.," *Magn. Reson. Med.*, vol. 42, no. 6, pp. 1072–81, Dec. 1999.
- [83] I. Diaz, P. Boulanger, R. Greiner, and A. Murtha, "A critical review of the effects of denoising algorithms on MRI brain tumor segmentation.," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2011, pp. 3934–7, Jan. 2011.
- [84] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, "MRI denoising using non-local means," *Med. Image Anal.*, vol. 12, pp. 514–523, 2008.
- [85] C. Gaser and P. Coupé, "Impact of non-local means filtering on brain tissue segmentation," *Organ. Hum. Brain Mapp. 2010 Annu. Meet. United States*, vol. 4, 2010.
- [86] J. V. Manjón, P. Coupé, A. Buades, D. Louis Collins, and M. Robles, "New methods for MRI denoising based on sparseness and self-similarity.," *Med. Image Anal.*, vol. 16, no. 1, pp. 18–27, Jan. 2012.
- [87] B. Fischl, D. Salat, E. Busa, and M. Albert, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, pp. 341–355, 2002.
- [88] "Morphometry Protocols," 2009. [Online]. Available: http://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki?action=AttachFile&do=get&target=FreeSurfer_Suggested_Morphometry_Protocols.pdf. [Accessed: 01-Oct-2013].
- [89] X. Han and B. Fischl, "Atlas renormalization for improved brain MR image segmentation across scanner platforms," *IEEE Trans. Med. Imaging*, vol. 26, no. 4, pp. 479–486, 2007.

- [90] B. B. Avants, N. J. Tustison, J. Wu, P. a Cook, and J. C. Gee, "An open source multivariate framework for n-tissue segmentation with evaluation on public data.," *Neuroinformatics*, vol. 9, no. 4, pp. 381–400, Dec. 2011.
- [91] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage*, vol. 17, pp. 825–841, 2002.
- [92] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, pp. 143–156, 2001.
- [93] S. Klein, M. Staring, and K. Murphy, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 196–205, 2010.
- [94] Wise.io, "Benchmarking Random Forest Classification." [Online]. Available: <http://about.wise.io/blog/2013/07/15/benchmarking-random-forest-part-1/>. [Accessed: 05-Mar-2013].
- [95] M. Dubuisson and A. Jain, "A modified Hausdorff distance for object matching," *Proc. Int. Conf. Pattern Recognition, Jerusalem*, pp. 566–568, 1994.