

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR
CENTRO DE CIÊNCIA EXATAS E DE TECNOLOGIA– CCET
DEPARTAMENTO DE COMPUTAÇÃO– DC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO– PPGCC

Katia Maria Poloni

**Detecção e classificação automáticas de
alterações estruturais cerebrais em
imagens de ressonância magnética para
o auxílio ao diagnóstico do Alzheimer**

São Carlos
Setembro de 2021

Katia Maria Poloni

**Detecção e classificação automáticas de
alterações estruturais cerebrais em
imagens de ressonância magnética para
o auxílio ao diagnóstico do Alzheimer**

Tese apresentada ao Programa de pós-graduação em ciência da computação do Centro de Ciência exatas e de tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutora em Ciência da Computação.

Área de concentração: Processamento de Imagens e Sinais

Orientador: Prof. Dr. Ricardo José Ferrari

São Carlos
Setembro de 2021



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Defesa de Tese de Doutorado da candidata Katia Maria Poloni, realizada em 21/09/2021.

Comissão Julgadora:

Prof. Dr. Ricardo José Ferrari (UFSCar)

Prof. Dr. Cesar Henrique Comin (UFSCar)

Prof. Dr. Alexandre Luis Magalhães Levada (UFSCar)

Prof. Dr. Hélio Pedrini (UNICAMP)

Prof. Dr. Leandro Alves Neves (UNESP)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Computação.

*Este trabalho é dedicado à todos que sempre estiveram ao meu lado, meus amigos
e minha família.*

Agradecimentos

Em primeiro lugar, eu agradeço a Deus por toda a proteção que existe ao meu redor. Agradeço também aos meus pais, Rui e Nadir, por terem me dado a vida, amor, apoio e dedicação em todas as etapas da minha vida. Agradeço às minhas irmãs, Nadia e Flávia, por toda a parceria de vida. Ao meu noivo e companheiro, Jonas Felinto, por sempre estar ao meu lado, pela cumplicidade, dedicação e amor incondicional. Agradeço às minhas “filhas” Duda e Luna, por serem a minha alegria diária e mudarem a minha vida. Aos meus cunhados, João e Clara, e aos meus sogros, Alan e Gizilene, por me apoiarem e por todos os bons momentos que pudemos compartilhar.

Agradeço ao meu orientador Prof. Dr. Ricardo José Ferrari, por sua orientação, compartilhar seus conhecimentos, me apoiar, estar presente e sempre me motivar a melhorar. Agradeço também a todos os meus amigos (do BIP-Group, do laboratório Gapis e da vida), por dividirem momentos de descontração e por todas as discussões construtivas compartilhadas durante o desenvolvimento desta pesquisa. Por fim, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) — Código Financeiro 001 e à Fundação de Pesquisa de São Paulo (FAPESP) — número de concessão 2018/06049-5, por financiarem esta pesquisa.

Andar com fé eu vou, que a fé não costuma “faiá.”

Gilberto Gil.

Resumo

A doença de Alzheimer é uma condição neurodegenerativa progressiva e irreversível com desenvolvimento caracterizado por atrofia cerebral com desenvolvimento assimétrico. Ao longo da vida, um cérebro cognitivamente saudável apresenta mudanças estruturais, no entanto, com o desenvolvimento do Alzheimer, essas mudanças são mais aceleradas, com menor intensidade no estágio de comprometimento cognitivo leve, e mais acentuada a medida que a doença progride. Neste projeto de pesquisa foram desenvolvidas três abordagens para auxiliar o diagnóstico de pacientes com comprometimento cognitivo leve e doença de Alzheimer leve, sendo que cada uma analisou propriedades distintas e complementares relacionadas à doença, a nomear a assimetria, a atrofia e o envelhecimento biológico cerebral, por fim, as abordagens foram combinadas para medir o impacto nos resultados. Na análise de assimetrias foram extraídos atributos multi-escala das regiões hipocampais e criado de um índice de assimetria. Os resultados de classificação foram comparáveis a outros estudos focados nas regiões hipocampais e o índice de assimetria criado apresentou resultados estatisticamente significativos e consistentes com a literatura médica. Na análise de atrofia, pontos salientes discriminativos entre as populações foram automaticamente selecionados e posteriormente classificados. A seguir, as imagens foram classificadas com base em uma avaliação quantitativa obtida a partir dos resultados dos pontos. Os resultados de classificação superaram muitos estudos publicados na literatura e são comparáveis a outros. Na análise do envelhecimento biológico do cérebro foi desenvolvido um modelo de estimação etária utilizando aprendizado profundo e imagens de pacientes cognitivamente normais com a mesma faixa etária de pacientes com declínio cognitivo. O erro de estimação etária entre os grupos apresentou resultados estatisticamente significativos, além de apresentar correlação significativa com o teste clínico *Mini-Mental State Examination*. Os resultados foram competitivos entre os estudos já existentes. Por fim, a combinação das duas primeiras abordagens trouxeram contribuições positivas para os resultados, com ganhos de até 3% na AUC.

Palavras-chave: doença de Alzheimer, atlas probabilístico, detecção de pontos salientes, atrofia estruturais cerebrais, assimetrias hipocampais, idade biológica do cérebro, redes neurais convolucionais, classificação de imagens de ressonância magnética.

Abstract

Alzheimer’s disease is a progressive and irreversible neurodegenerative condition with development characterized by asymmetrical brain atrophies. Across life, due to neuronal aging, a cognitively healthy brain presents structural changes. However, with the development of the Alzheimer’s, these changes are more accentuated, being less intensive in the prodromal stage of Alzheimer’s, known as mild cognitive impairment, and more pronounced as the disease progresses. In this research, we developed three approaches to aid in diagnosing patients with mild cognitive impairment and mild Alzheimer’s disease, each of which analyzed distinct and complementary properties related to the disease, namely asymmetry, atrophy, and biological brain aging. In addition, we combined and tested the approaches to measure impact on outcomes. In the analysis of structural asymmetries, we extracted multi-scale attributes of the hippocampal regions and created an asymmetry index. We obtained classification results comparable to other studies focused on the hippocampal regions, and the created asymmetry index showed statistically significant results consistent with the medical literature. In the analysis of structural atrophies, we automatically selected and classify discriminative landmark points between populations. Next, we classify the images based on a quantitative assessment obtained from the results of the landmark points. The obtained image classification results surpassed many studies published in the literature and are comparable to others. In the analysis of the biological aging of the brain, we developed an age estimation model using deep learning and images of cognitively normal patients with the same age range as patients with mild cognitive impairment and mild Alzheimer’s disease. The age estimation error between the groups showed statistically significant results and presented a significant correlation with the Mini-Mental State Examination clinical test. The estimation results were competitive among existing studies, but the classification results were below expectations. Finally, combining the first two approaches brought positive contributions to the results, with up to 3% gains in AUC.

Keywords: Alzheimer’s disease, probabilistic atlases, salient points detection, structural brain atrophies, hippocampal asymmetry, brain-age estimation, convolutional neural networks, MR image classification.

Lista de ilustrações

Figura 1 – DA <i>continuum</i> . Extraído, editado e traduzido de Association (2020).	41
Figura 2 – Diferenças entre um cérebro saudável e com a doença de Alzheimer. Extraído de (ASSOCIATION, 2020).	42
Figura 3 – Representação visual das respostas dos filtros na vista sagital.	50
Figura 4 – Representação gráfica do espaço log-esférico ao redor de um ponto saliente para o cômputo do descritor. (a) <i>shape context</i> 3D global; (b) <i>shape context</i> 3D local.	54
Figura 5 – Política de decaimento da CLR.	67
Figura 6 – Visão geral das principais abordagens desenvolvidas nesta pesquisa.	98
Figura 7 – Vista sagital de uma imagem de RM com ponderação em T1- <i>w</i> (a) antes e (b) após aplicação do NLM.	101
Figura 8 – Vista sagital de uma imagem de RM com ponderação em T1- <i>w</i> (a) antes, (b) após a aplicação do N4-ITK e (c) <i>bias field</i> removido.	102
Figura 9 – (a) Histograma da imagem de referência; (b) histograma de uma imagem de estudo antes e (c) após a aplicação do algoritmo.	103
Figura 10 – Comparação na vista sagital entre o “tabuleiro de xadrez” formado por uma imagem de estudo e a imagem da NAC (a) antes e (b) após o correregistro afim.	104

Figura 11 – Vista coronal de uma imagem de RM com a máscara do encéfalo obtida pelo ROBEX sobreposta à imagem original.	104
Figura 12 – Fluxograma da sequência de etapas utilizadas no desenvolvimento da análise de atrofia estruturais.	105
Figura 13 – Fluxograma para a seleção dos pontos salientes do atlas.	107
Figura 14 – Mapas probabilísticos das imagens de RM utilizados como fonte de informação para a extração dos atributos: (a) WM, (b) GM, e (c) CSF.	108
Figura 15 – Fluxograma da sequência de etapas utilizadas no desenvolvimento da análise de assimetrias estruturais.	110
Figura 16 – Vista coronal de uma imagem de RM com detecção do MSP, em (a) um exemplo da detecção, (b) o hemisfério direito e (c) o hemisfério esquerdo rotacionado.	111
Figura 17 – Ilustração da assimetria hemisférica hipocampal utilizando diagramas de rosas residuais obtidos de pacientes pertencentes aos grupos CN, CCL, DA.	113
Figura 18 – Fluxograma da sequência de etapas utilizada para o desenvolvimento desta abordagem.	117
Figura 19 – Arquitetura da <i>EfficientNet</i> B0 usada neste trabalho.	119
Figura 20 – Exemplo de dezesseis imagens aleatórias de um <i>batch</i>	119
Figura 21 – A distribuição de idade da população de estudo do conjunto de dados originais e do conjunto de dados aumentados. (a) Para o experimento com idade entre 20 e 70 anos, e (b) para o experimento maior que 70 anos.	120
Figura 22 – Divisão dos <i>folds</i> utilizada para a metodologia de dois níveis de classificação.	126
Figura 23 – Representação visual (vista coronal) da imagem em (a) T1- <i>w</i> e do (b) maior e (c) menor autovalor do cérebro resultantes da congruência fase.	130
Figura 24 – Sequência de etapas utilizadas para a avaliação dos parâmetros dos descritores.	131
Figura 25 – Influência dos parâmetros do descritor no casamento entre os pontos salientes.	132

Figura 26 – Influência dos parâmetros do descritor na classificação dos pontos salientes.	133
Figura 27 – Atlas probabilístico de pontos salientes 3D da região hipocampal esquerda.	134
Figura 28 – Curvas ROC e valores de AUC correspondentes aos resultados de classificação de CN×CCL, CCL×DA e CN×DA.	139
Figura 29 – Gráficos de dispersão de quatro atributos (de um total de dez) correspondentes a um <i>fold</i> de cada um dos experimentos: CN×CCL, CCL×DA, e CN×DA, respectivamente.	140
Figura 30 – Maiores e menores probabilidades de predição para o experimento CN×CCL: (a) Melhor para CCL, (b) melhor para CN, (c) pior para CCL e (d) pior para CN.	141
Figura 31 – Maiores e menores probabilidades de predição para o experimento CCL×DA: (a) Melhor para CCL, (b) melhor para DA, (c) pior para CCL e (d) pior para DA.	142
Figura 32 – Maiores e menores probabilidades de predição para o experimento CN×DA: (a) Melhor para DA, (b) melhor para CN, (c) pior para DA e (d) pior para CN.	143
Figura 33 – Banco de filtros log-Gabor 3D utilizados nesta etapa.	146
Figura 34 – Curvas ROC e valores de AUC correspondentes aos resultados de classificação de (a) CN×CCL, (b) CCL×DA e (c) CN×DA.	149
Figura 35 – Boxplots das diferenças entre as médias de <i>IA</i> da variância por (a) grupos de diagnóstico e (b) gênero.	152
Figura 36 – A distribuição de idade da população para as imagens de validação e teste.	154
Figura 37 – Análise da faixa ótima da TA.	154
Figura 38 – Curvas da função de perda do modelo com e sem aumento de dados.	156
Figura 39 – Gráficos de correlação das idades biológicas estimadas com as idades cronológicas do modelo CNN para (a) o hipocampo esquerdo, (b) o hipocampo direito e (c) o <i>ensemble</i>	158
Figura 40 – Imagens de pacientes com idade cronológica de ≈ 43 anos.	159
Figura 41 – <i>Boxplots</i> dos valores $\Delta_{\text{Idade cérebro}}$ por grupo de diagnóstico.	161

Figura 42 – Gráfico de correlação do $\Delta_{\text{Idade c393erebro}}$ e MMSE.	162
Figura 43 – Imagens de diferena entre pacientes CN e DA, CN e CCL, e CCL e DA para os hipocampos esquerdo e direito nas vistas axial, coronal e sagital, respectivamente.	163
Figura 44 – Imagens de diferena entre as imagens m393dias de cada grupo de diagn393stico divididas por faixas et393rias entre 70 e 77 anos e entre 76 e 85 anos: o grupo CN, CCL e DA para o hipocampo esquerdo.	163
Figura 45 – Gr393fico de barras para a compara393o dos resultados para o experimento CN×CCL	167
Figura 46 – Gr393fico de barras para a compara393o dos resultados para o experimento CCL×DA	168
Figura 47 – Gr393fico de barras para a compara393o dos resultados para o experimento CN×DA	169
Figura 48 – Curvas ROC e valores de AUC correspondentes aos resultados de classifica393o de (a) CN×CCL , (b) CCL×DA e (c) CN×DA para os m393todos de atrofia, assimetria, uni393o das abordagens e ROI-based.	172

Lista de tabelas

Tabela 1 – Matriz de Confusão	68
Tabela 2 – Resumo dos principais trabalhos publicados na literatura relacionados ao tema desta pesquisa.	79
Tabela 3 – Resumo dos principais trabalhos publicados na literatura que analisam a assimetria estrutural.	83
Tabela 4 – Resumo dos principais trabalhos publicados na literatura para a predição da idade biológica do cérebro.	93
Tabela 5 – Conjunto de imagens da base de dados ADNI.	100
Tabela 6 – Conjunto de imagens da base de dados IXI.	100
Tabela 7 – Quantidade de imagens (apenas pacientes CN) utilizadas em cada etapa de treinamento.	117
Tabela 8 – Resultados da AUC média \pm desvio-padrão para os experimentos CN \times CCL, CCL \times DA, e CN \times DA realizados no nível de pontos salientes.	136
Tabela 9 – Resultados de classificação dos 10 <i>folds</i> de validação cruzada (média \pm desvio-padrão) para o experimento CN \times CCL.	137
Tabela 10 – Resultados de classificação dos 10 <i>folds</i> de validação cruzada (média \pm desvio-padrão) para o experimento CCL \times DA.	138
Tabela 11 – Resultados de classificação dos 10 <i>folds</i> de validação cruzada (média \pm desvio-padrão) para o experimento CN \times DA.	138

Tabela 12 – Resultados da classificação das imagens utilizando <i>nested cross-validation</i> com 10 <i>folds</i> (média \pm desvio) para cada medida estatística.	147
Tabela 13 – Resultados da classificação das imagens utilizando <i>nested cross-validation</i> com 10 <i>folds</i> (média \pm desvio) para a concatenação de duas medidas estatísticas	148
Tabela 14 – Resultados da classificação das imagens utilizando <i>nested cross-validation</i> com 10 <i>folds</i> (média \pm desvio) para a concatenação de três e quatro medidas estatísticas	148
Tabela 15 – Média e desvio padrão dos valores de <i>IA</i> por grupos de diagnóstico, resultados da ANOVA ($\alpha = 0,05$) e conclusões dos testes <i>post-hoc</i> HSD ($\alpha_t < 0,05$).	151
Tabela 16 – Média e desvio padrão dos valores de <i>IA</i> por gênero e resultados da ANOVA ($\alpha = 0,05$).	152
Tabela 17 – Hiperparâmetros usados nos experimentos.	155
Tabela 18 – Resultados da predição da idade do cérebro para o primeiro experimento.	157
Tabela 19 – Resultados da predição da idade do cérebro para o segundo experimento.	160
Tabela 20 – Resultados da predição da idade do cérebro para os grupos com CCL e DA.	160
Tabela 21 – Média e desvio padrão dos valores do $\Delta_{\text{Idade cérebro}}$ por grupos de diagnóstico, resultados da ANOVA ($\alpha = 0,05$) e conclusões dos testes <i>post-hoc</i> HSD ($\alpha_t < 0,05$).	161
Tabela 22 – Resultados da classificação das imagens utilizando <i>nested cross-validation</i> com 10 <i>folds</i> (média \pm desvio).	164
Tabela 23 – Resultados da classificação das imagens utilizando <i>nested cross-validation</i> com 10 <i>folds</i> (média \pm desvio) para a união das abordagens.	166
Tabela 24 – Resultados da classificação das imagens utilizando <i>nested cross-validation</i> com 10 <i>folds</i> (média \pm desvio) para cada abordagem.	170

Tabela 25 – Comparação indireta de nossos resultados sobre a classificação de imagens de RM de casos CN×CCL, CCL×CN e CN×AD com estudos relacionados. 173

Lista de siglas

Acc	Acurácia
ADNI	Alzheimer's Disease Neuroimaging Initiative
AUC	Área sob a curva - Area Under Curve
CCL	Comprometimento Cognitivo Leve
CLR	Taxa de aprendizagem cíclica - Cyclical learning rate
CN	Cognitivamente normal
CNNs	Convolutional neural networks
CSF	Líquido cefalorraquidiano - Cerebrospinal fluid
DA	Doença de Alzheimer
DTI	Diffusion Tensor Imaging
EM	Expectation Maximization
Espec	Especificidade
F1	Medida F1

FLOPS	Operações de ponto flutuante por segundo - Floating point operations per second
GH	Guy's Hospital
GM	Substância cinzenta - Gray Matter
GPR	Regressão do processo gaussiano - Gaussian process regression
HH	Hammersmith Hospital
HMRF	Hidden Markov Random Field
HOG	Histograma de gradientes orientados - Histogram of oriented gradients
HPF	Fração do parênquima do hipocampo - Hippocampal Parenchymal Fraction
IA	Índice de Assimetria
IDA	Image & Data Archive
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IOP	Institute of Psychiatry
IXI	Information eXtraction from Images
LDA	Linear discriminant analysis
LONI	Laboratory of Neuro Imaging's
MAE	Mean absolute error
MBCConv	Mobile inverted bottleneck convolution
ME	Mean error
MMSE	Mini-Mental State Examination
MNI	Montreal Neurological Institute

MPRAGE	Magnetization Prepared Rapid Gradient Echo
MPW	Manifold Parzen Windows
MSP	Plano sagital médio - MidSagittal Plane
N3	Nonparametric Nonuniform intensity Normalization
NAC	Neuroimage Analysis Center
NLM	NonLocal Means
OMS	Organização Mundial da Saúde
PC	Congruência de fase - Phase Congruency
PCA	Análise dos componentes principais - Principal component analysis
pCCL	Comprometimento Cognitivo Leve progressivo
pdf	Função densidade probabilidade - Probability density function
r	Coefficiente de correlação de Pearson
RBF	Função de base radial - Radial basis function
RF	RadioFrequência
RM	Ressonância Magnética
RMSE	Root mean squared error
ROBEX	Robust Brain Extraction
ROI	Região de interesse - Region of Interest
sCCL	Comprometimento Cognitivo Leve estável (stable)
Sens	Sensibilidade
SNIPe	Scoring by nonlocal image patch estimator
TA	Taxa de aprendizado

VBM	Morfometria baseada em voxel - Voxel based Morphometry
WLS	Weighted Least Squares
WM	Substância branca - White matter

Sumário

1	INTRODUÇÃO	29
1.1	Contexto e motivação	29
1.2	Aspectos de originalidade	31
1.3	Questões de pesquisa	33
1.4	Objetivos	34
1.4.1	Objetivo geral	34
1.4.2	Objetivos específicos	34
1.5	Contribuições alcançadas e publicações	35
1.6	Organização do trabalho	37
2	CONCEITOS RELACIONADOS	39
2.1	Doença de Alzheimer	39
2.2	Ressonância Magnética	42
2.3	Algoritmos de pré-processamento	43
2.3.1	Redução de ruído	44
2.3.2	Correção de <i>bias field</i>	45
2.3.3	Padronização de histogramas	45
2.3.4	Corregistro das imagens	46
2.3.5	Extração da máscara do encéfalo	46
2.4	Filtros Log-Gabor 3D	47

2.5	Momentos de segunda ordem	48
2.6	Filtragem direcional	49
2.7	Congruência de fase 3D	49
2.7.1	Detecção de pontos salientes	52
2.7.2	Descritores	52
2.7.3	Casamento	54
2.7.4	Detecção do plano sagital médio	55
2.8	Atlas probabilístico de pontos salientes 3D	55
2.9	Segmentação de tecidos	56
2.10	Classificador SVM	57
2.11	Testes estatísticos	58
2.11.1	Welch <i>test two-tailed</i>	59
2.11.2	ANOVA	60
2.11.3	Tukey HSD	61
2.12	Redes neurais convolucionais	62
2.12.1	Arquiteturas	62
2.12.2	Taxa de aprendizagem cíclica	67
2.13	Métricas de avaliação	68
2.13.1	Classificação	68
2.13.2	Regressão	69
2.14	Considerações finais	70
3	REVISÃO BIBLIOGRÁFICA	71
3.1	Atrofias estruturais	71
3.1.1	Trabalhos correlatos	73
3.2	Assimetrias estruturais	80
3.2.1	Trabalhos correlatos	80
3.3	Predição da idade biológica do cérebro	84
3.3.1	Trabalhos correlatos	85
3.4	Considerações finais	94
4	MATERIAIS E MÉTODOS	97
4.1	Visão geral	97

4.2	Base de imagens	98
4.2.1	<i>NAC Brain Atlas</i>	98
4.2.2	ADNI	99
4.2.3	IXI	99
4.3	Pré-processamento	100
4.3.1	Parâmetros e análises qualitativas	101
4.4	Análise de atrofas estruturais	105
4.4.1	Definição das regiões de interesse	106
4.4.2	Seleção de pontos salientes	106
4.4.3	Extração de atributos dos pontos salientes	107
4.4.4	Classificação das imagens	108
4.5	Análise de assimetrias estruturais	109
4.5.1	Plano sagital médio	110
4.5.2	Filtragem direcional	111
4.5.3	Análise entre hemisférios	112
4.5.4	Classificação das imagens	114
4.5.5	Análise estatística dos atributos	114
4.6	Análise do erro de predição da idade biológica do cérebro	116
4.6.1	Base de imagens	116
4.6.2	Definição das regiões de interesse	117
4.6.3	CNN	118
4.6.4	Aumentos dos dados	118
4.6.5	Análise estatística	120
4.6.6	Classificação	121
4.7	União das abordagens	121
4.8	Comparação dos resultados	122
4.9	Considerações finais	123
5	RESULTADOS E DISCUSSÕES	125
5.1	Análise de atrofas estruturais	125
5.1.1	Configurações experimentais	125
5.1.2	Análise dos parâmetros	128
5.1.3	Resultados	135

5.2	Análise de assimetrias estruturais	143
5.2.1	Configurações experimentais	143
5.2.2	Análise dos parâmetros	144
5.2.3	Resultados	145
5.3	Análise do erro de predição da idade biológica do cérebro	153
5.3.1	Configurações experimentais	153
5.3.2	Resultados	155
5.4	União das abordagens	165
5.5	Comparação dos resultados	169
5.5.1	Comparação direta	170
5.5.2	Comparação indireta	172
5.6	Considerações finais	173
6	CONCLUSÕES	175
6.1	Análise de atrofas estruturais	175
6.2	Análise de assimetrias estruturais	177
6.3	Análise do erro de predição da idade biológica do cérebro	178
6.4	União das abordagens	179
6.5	Considerações finais	180
	REFERÊNCIAS	181

Capítulo 1

Introdução

Este capítulo apresenta o contexto no qual a pesquisa está inserida, a motivação para fornecer soluções para o diagnóstico da doença de Alzheimer, os objetivos, produções científicas geradas e a estrutura de organização do texto.

1.1 Contexto e motivação

A doença de Alzheimer (DA) é a causa mais comum de demência, representando cerca de 60% a 80% de todos os casos. É uma doença neurodegenerativa, progressiva e irreversível que afeta predominantemente a população idosa. Geralmente se desenvolve lentamente, apresentando piora gradual da função cerebral à medida que as células morrem (ASSOCIATION, 2020).

Visto que o fator de risco mais significativo para o desenvolvimento da DA é a idade avançada, o crescente envelhecimento da população alarma os órgãos de saúde. Segundo a Organização Mundial da Saúde (OMS), a população global com Alzheimer aproxima-se dos 35 milhões e estima-se que até 2050 a quantidade pode triplicar (ALZHEIMER'S DISEASE INTERNATIONAL, 2020), além disso, estima-se um gasto mundial com demências de 818 bilhões de dólares. Existe um alto ônus tanto financeiro quanto social, à sociedade e aos indivíduos, com o

desenvolvimento dessas doenças (COLE; FRANKE, 2017). No Brasil, a população idosa chega aos 15 milhões, sendo que cerca de 1,2 milhão de pessoas possuem a DA. Devido à gravidade do cenário, em 2012, a OMS incluiu a demência como uma condição prioritária de saúde (ZHANG et al., 2017; ASSOCIATION, 2020).

O diagnóstico clínico da DA é um processo desafiador de avaliação, que segue os critérios clínicos estabelecidos pelo *National Institute of Neurological and Communicative Disorders and Stroke e Alzheimer's Disease and Related Disorders Association* (NINCDS—ADRDA) (BEACH et al., 2012) e que requer a eliminação de outras causas potenciais de demência (HYMAN et al., 2012; MCKHANN et al., 2011). Alguns estudos (LIM et al., 1999; KLATKA et al., 1996; BEACH et al., 2012) avaliaram previamente a precisão do diagnóstico clínico da DA em comparação com resultados de autópsias e reportaram taxas de imprecisão entre 12% e 23%, considerando os pacientes diagnosticados com DA que não possuíam patologias suficientes na autópsia capazes de explicar a presença da demência (GAUGLER et al., 2013).

O imageamento por ressonância magnética (RM) estrutural desempenha um papel cada vez mais significativo em ensaios clínicos e estudos relacionados à DA. O principal motivo é sua alta sensibilidade na detecção de alterações cerebrais estruturais sutis (JOHNSON et al., 2012), que resultam em imagens com contraste superior e detalhes anatômicos mais finos. Essa propriedade torna possível detectar alterações anatômicas induzidas por doenças no cérebro (LUO; TANG, 2017) e ajudar no diagnóstico de DA antes mesmo do aparecimento de sintomas clínicos.

O cérebro humano exhibe um processo biológico complexo de mudanças ao longo da vida (COLE et al., 2017), caracterizado por padrões regionais específicos e não lineares de eventos coordenados e sequenciais durante o desenvolvimento (CHERUBINI et al., 2016), com um declínio geral no desempenho cognitivo e atrofia cerebral. No entanto, na presença de condições patológicas, essas mudanças são geralmente mais severas.

Durante o desenvolvimento da DA, a atrofia é generalizada e se desenvolve de maneira assimétrica, sendo frequentemente caracterizada pela atrofia regional desbalanceada da substância cinzenta (*Gray Matter* — GM) do cérebro, cujo processo de degradação afeta primeiro, e mais severamente, as regiões do hemisfério esquerdo (TOGA; THOMPSON, 2003) — sendo o hipocampo uma das primeiras

estruturas cerebrais a experienciar ambos eventos com maior severidade (GUADALUPE et al., 2014; WOOLARD; HECKERS, 2012). Posteriormente, outras partes do cérebro também são afetadas e começam a diminuir à medida que mais neurônios morrem. Na fase final da doença, o dano é generalizado e o tecido cerebral encolhe significativamente (ABBOTT, 2011; ORGANIZATION; INTERNATIONAL, 2012).

Apesar de décadas de pesquisas, ainda não existe um teste definitivo para diagnosticar a DA, sendo o diagnóstico definitivo só podendo ser efetuado após a morte do paciente para a realização da confirmação histopatológica, indicando a presença excessiva da proteína beta amiloide e emaranhados neurofibrilares no cérebro (VEMURI; JACK, 2010). Os trabalhos voltados à detecção da DA tem como foco principal a detecção da doença em seu estágio inicial, chamado comprometimento cognitivo leve (CCL), pois pode auxiliar no desenvolvimento de novas terapias e melhorar o tratamento. No entanto, devido ao diagnóstico clínico para provável DA ser impreciso, a precisão diagnóstica para a classe CCL é ainda menor (SHAW et al., 2009), visto que esta representa um estágio de transição entre o envelhecimento normal e uma síndrome demencial, com sintomas mais difíceis de serem identificados e alterações anatômicas mais sutis.

1.2 Aspectos de originalidade

Um dos principais enfoques desta pesquisa consiste em analisar atrofia estruturais em imagens RM para auxiliar no diagnóstico da DA leve e do CCL. Até o momento, muitos trabalhos (SØRENSEN et al., 2016; AHMED et al., 2015b; AHMED et al., 2015a; COUPÉ et al., 2012b; ZHANG et al., 2016; ZHANG; LIU; SHEN, 2017; ZHANG et al., 2017) têm sido publicados na literatura com o mesmo objetivo de classificar as imagens nos grupos cognitivamente normais (CN), CCL e DA. Em geral, esses trabalhos procuraram identificar alterações nos tecidos cerebrais usando técnicas baseados em *voxel*, regiões de interesse (*Region of interest* — ROIs), *patches* e aprendizagem profunda, no entanto, poucas análises foram realizadas em relação aos atributos extraídos, pois todos foram comumente considerados valiosos e usados para classificação, embora, nem todos representem alterações cerebrais causadas por patologias da DA ou do CCL. Dessa maneira, este

estudo visa realizar a classificação das imagens utilizando atributos que representam alterações cerebrais causadas por patologias desses estágios, unindo técnicas baseadas em ROIs e pontos salientes.

Outro aspecto importante deste trabalho é analisar a assimetria estrutural hipocampal, visto que tal informação foi relatada como parte do desenvolvimento da DA. Nos últimos anos, estudos avaliaram assimetrias hemisféricas em imagens de RM buscando examinar diferentes alterações decorrentes do envelhecimento saudável (ASHBURNER et al., 1998; GOOD et al., 2001; LUDERS et al., 2004; LUCARELLI et al., 2013; KOVALEV; KRUGGEL; CRAMON, 2002; HERVÉ et al., 2006) e doenças cognitivas (HERBERT et al., 2005; ORTIZ et al., 2019; MARTINS et al., 2019; SARICA et al., 2018; ARDEKANI et al., 2019; BARNES et al., 2005; THOMPSON et al., 2003). Dentre esses estudos, os que avaliaram as assimetrias na DA (SARICA et al., 2018; ARDEKANI et al., 2019; BARNES et al., 2005; THOMPSON et al., 2003) geralmente se concentraram em diferenças volumétricas entre os hipocampus direito e esquerdo. No entanto, nenhum deles submeteram seus atributos para classificação, sendo, nesses casos, realizados apenas testes estatísticos de avaliação. Por esse motivo, este trabalho propôs a criação de uma técnica para realizar a classificação de imagens de RM usando apenas atributos que apresentem diferenças significativas entre os grupos de diagnóstico (CN, CCL e DA leve) quando testados para assimetria, referenciados ao longo do texto como atributos assimétricos. Ao invés de utilizar o volume hipocampal, propusemos atributos obtidos das respostas da filtragem direcional multiescalas das regiões hipocampais, os quais foram comparados entre os hemisférios cerebrais, direito e esquerdo.

Este trabalho também propoe a estimação da idade biológica do cérebro de pacientes com CCL e DA leve. A previsão da idade do cérebro a partir de imagens de RM e usando técnicas de aprendizado de máquina, tem sido amplamente estudada e tem fornecido cada vez mais percepções sobre os efeitos de alterações cerebrais associadas à idade e como as doenças afetam o envelhecimento do cérebro (COLE; FRANKE, 2017). Diferentes métodos foram propostos para esta tarefa, com a utilização de engenharia de atributos (FRANKE et al., 2010; FUJIMOTO et al., 2017; MADAN; KENSINGER, 2018; PARDAKHTI; SAJEDI, 2017; SU et al., 2011) e aprendizagem profunda (DINSDALE et al., 2021; COLE et al., 2017;

PENG et al., 2021; HUANG et al., 2017; ITO et al., 2018; UEDA et al., 2019). Em geral, a estimaco etria do crebro  realizada em uma faixa etria bastante extensa (\approx entre 20 e 80 anos) e desbalanceada, devido  maior disponibilidade de imagens. No entanto, o uso de um grande intervalo etrio com desbalanceamento dificulta a anlise comparativa sem vis entre o envelhecimento saudvel do crebro e doenas especficas. Sendo assim, nesta pesquisa optou-se por criar uma metodologia utilizando aprendizado profundo para prever a idade biolgica do crebro de indivduos CN com faixa etria maior que setenta anos para avaliar as diferenas entre a predico da idade do crebro de pacientes com DA leve e CCL. Por fim, este trabalho props unir os resultados obtidos pelas trs abordagens desenvolvidas nesta pesquisa para verificar se so complementares e contribuem positivamente para melhorar os resultados individuais obtidos.

1.3 Questes de pesquisa

Sabe-se que o desenvolvimento da DA consiste em um processo bastante complexo, em que so analisados mltiplos fatores para auxiliar no diagnstico da doena. Dentre estes fatores, esto os que fazem uso de imagens de RM, em que as imagens so analisadas com base em comparaes com imagens de crebros considerados cognitivamente saudveis. Sendo assim, as questes investigadas nesta Tese, tiveram enfoque em explorar essas comparaes de maneira a analisar conceitos que so distintos e se complementam durante o desenvolvimento da DA, como a atrofia, a assimetria e a idade do crebro. Seguindo os aspectos de originalidade propostos nesta pesquisa, foram formulados os seguintes questionamentos:

1.  possvel construir vetores de atributos de atrofia estrutural hipocampal interpretveis, com baixa dimensionalidade e que representem alteraes cerebrais causadas por patologias da DA ou do CCL que ainda sejam representativos entre os grupos?
2.  possvel que atributos extrados apenas com metodologias que englobem a assimetria estrutural hipocampal e que apresentem diferenas estatsticas significativas entre os grupos comparados, sejam descritivos e discriminativos para classificar as imagens de RM?

3. É possível que um modelo treinado para prever a idade do cérebro de pacientes CN obtenha diferenças significativas no erro de estimação ao ser comparado com imagens do cérebro de pacientes com declínio cognitivo e estes sejam também utilizados para classificar as imagens de RM?
4. Quais dos três conceitos possuem maior poder predito em termos de classificação?
5. Visto que os conceitos são complementares, a união de métodos/modelos construídos com os mesmos é capaz de melhorar os resultados obtidos pelo estado da arte?

No decorrer do texto desta Tese, todos os questionamentos são respondidos.

1.4 Objetivos

1.4.1 Objetivo geral

Com base no que foi exposto, o principal objetivo desta pesquisa foi estudar e desenvolver técnicas computacionais automáticas capazes de classificar alterações estruturais cerebrais em imagens de RM para auxiliar o diagnóstico de pacientes com CCL e DA leve. Para alcançar tal objetivo, foram propostas três abordagens restritas às regiões hipocampais. Cada abordagem buscou analisar os conceitos introduzidos na seção anterior, a nomear a atrofia e a assimetria, e a análise do envelhecimento biológico do cérebro.

1.4.2 Objetivos específicos

Os objetivos específicos definidos para esta pesquisa foram:

- a) Adquirir, organizar e preprocessar as imagens clínicas de RM utilizando técnicas comumente utilizadas na literatura de processamento de imagens médicas;
- b) Estudar e desenvolver uma metodologia usando engenharia de atributos para detectar atrofia estrutural em imagens de RM para o auxílio ao

diagnóstico da DA utilizando atributos que representam alterações cerebrais causadas por patologias da DA leve ou do CCL;

- c) Estudar e desenvolver uma metodologia usando engenharia de atributos para classificar as imagens de RM usando apenas atributos que apresentam diferenças significativas entre os grupos de diagnóstico, CN, CCL e DA leve, quando testados para assimetria;
- d) Estudar e desenvolver uma metodologia para prever a idade do cérebro utilizando aprendizado profundo e avaliar as diferenças entre a predição da idade biológica do cérebro de pacientes CN que tenham a mesma faixa etária que pacientes com DA leve e CCL.

O texto desta Tese foi estruturada de maneira à explorar individualmente cada abordagem, e no fim, todas foram combinadas para avaliação.

1.5 Contribuições alcançadas e publicações

As contribuições científicas desta Tese consistiram na produção de algoritmos de código aberto e produções de artigos. É importante mencionar que a abordagem de atrofia estruturais possui conceitos semelhantes aos utilizados durante o Mestrado, cujos trabalhos foram publicados no início do doutorado (POLONI; FERRARI, 2018; POLONI et al., 2018). Além disso, devido à utilização de conceitos trabalhados durante este projeto, outras produções, nas quais tive coautoria, também foram geradas.

As produções que possuem relação direta com a pesquisa são:

- a) **Poloni, K. M.** and Ferrari, R. J. “*Detection and classification of hippocampal structural changes in MR images as biomarker for Alzheimer’s disease*”, 18th International Conference on Computational Science and its Applications, 2018 (POLONI; FERRARI, 2018).
- b) **Poloni, K. M.**, Villa-Pinto C. H., Souza, B. S. and Ferrari, R. J. “*Construction and application of a probabilistic atlas of 3D landmark points for*

initialization of hippocampus mesh models in brain MR images”, 18th International Conference on Computational Science and its Applications, 2018 (POLONI et al., 2018).

- c) Oliveira I. A. D., **Poloni, K. M.** and Ferrari, R. J. for the ADNI “*Exploring hippocampal asymmetrical features from magnetic resonance images for the classification of Alzheimer’s disease*”, IEEE 33rd International Symposium on Computer Based Medical Systems, 2020 (OLIVEIRA; POLONI; FERRARI, 2020).
- d) **Poloni, K. M.**, Oliveira, I. A. D., Tam, R. and Ferrari, R. J. for the ADNI “*Brain MR image classification for Alzheimer’s disease diagnosis using structural hippocampal asymmetrical attributes from directional 3-D log-Gabor filter responses*”, Neurocomputing (POLONI et al., 2021).
- e) Chaves Cambui, V. H., **Poloni, K. M.** and Ferrari, R. J., “*Classification of brain MR images for the diagnosis of Alzheimer’s disease based on features individually extracted from the three main brain tissues*”, Brazilian Technology Symposium (CAMBUI; POLONI; FERRARI, 2021).
- f) **Poloni, K. M.** and Ferrari, R. J., “*Automated detection, selection and classification of hippocampal landmark points for the diagnosis of Alzheimer’s disease*”, Computer Methods and Programs in Biomedicine (POLONI; FERRARI, 2021)
- g) **Poloni, K. M.** and Ferrari, R. J., “*A deep ensemble hippocampal CNN model for brain age estimation applied to Alzheimer’s diagnosis*”, Expert Systems with Applications (POLONI; FERRARI et al., 2022)

As produções que possuem relação indireta com a pesquisa são:

- a) Felinto, J. C., **Poloni, K. M.**, Freire, P. G. L., Aily, J. B. ; Almeida, A. C., Pedroso, M. G., Mattiello, S. M. and Ferrari, R. J., “*Automatic segmentation and quantification of thigh tissue in CT images*”, 18th International Conference on Computational Science and its Applications, 2018 (FELINTO et al., 2018)

- b) Souza, B. S. ; **Poloni, K. M.** and Ferrari, R. J. . “*Detector of 3-D salient points based on the dual-tree complex wavelet transform for the positioning of hippocampi meshes in magnetic resonance images*”, Journal of Neuroscience Methods, 2020 (SOUZA; POLONI; FERRARI, 2020).
- c) Moura Araujo, M. R., **Poloni, K. M.** and Ferrari, R. J., “*Assessment of linear and non-linear feature projections for the classification of 3-D MR images on cognitively normal, mild cognitive impairment and Alzheimer’s disease*”, 21st International Conference on Computational Science and its Applications, 2021 (ARAÚJO; POLONI; FERRARI, 2021).
- d) Frascá, F. N., **Poloni, K. M.** and Ferrari, R. J., “*Automatic Extraction of the Midsagittal Surface from T1-weighted MR Brain Images using a Multiscale Filtering Approach*”, 21st International Conference on Computational Science and its Applications, 2021 (FRASCÁ; POLONI; FERRARI, 2021).

1.6 Organização do trabalho

Este trabalho é composto por mais cinco capítulos. O conteúdo de cada um deles é o seguinte:

- a) Capítulo 2: contém os conceitos teóricos das técnicas utilizadas nesta pesquisa.
- b) Capítulo 3: aborda a revisão do estado da arte relacionando as diferentes abordagens e técnicas utilizadas.
- c) Capítulo 4: contém informações sobre as bases de dados utilizadas na pesquisa, bem como uma descrição das metodologias desenvolvidas.
- d) Capítulo 5: expõe os resultados obtidos e discussões resultantes.
- e) Capítulo 6: são expostas as conclusões obtidas e sugeridos trabalhos futuros.

Capítulo 2

Conceitos Relacionados

Este capítulo apresenta conceitos e a fundamentação teórica das técnicas utilizadas que são relevantes para o entendimento desta pesquisa. É apresentada uma descrição de aspectos fundamentais da doença de Alzheimer e também uma explicação resumida do processo de formação e aquisição das imagens de ressonância magnética. Em relação à fundamentação teórica, é apresentada uma descrição sobre os algoritmos de pré-processamento, o banco de filtros log-Gabor 3D, os métodos de congruência de fase 3D e atlas probabilístico de pontos salientes 3D utilizados para processar as imagens. Além disso, são apresentados detalhes do classificador SVM, arquitetura de redes neurais convolucionais e das técnicas estatísticas e métricas utilizadas para a validação dos resultados.

2.1 Doença de Alzheimer

Histologicamente, a DA está associada ao acúmulo excessivo da proteína beta amiloide, que se acumula nos espaços entre as células nervosas, e a deposição da proteína tau, que formam emaranhados neurofibrilares intracelulares (BASSIONY et al., 2015), que podem levar à morte dos neurônios. A destruição e morte das

células nervosas causa falha de memória, alterações de personalidade, problemas nas atividades diárias e outros sintomas da DA. Os cientistas ainda não sabem exatamente qual o papel das placas e emaranhados na doença, mas acreditam que, de alguma forma, eles desempenham um papel crítico no bloqueio da comunicação entre as células nervosas e na interrupção dos processos que elas precisam para sobreviver (ASSOCIATION, 2020). Estudos de autópsia mostram que a maioria das pessoas desenvolve algumas placas e emaranhados à medida que envelhece, no entanto, nas pessoas com DA, o desenvolvimento é mais acentuado e segue um padrão previsível, começando nas áreas do cérebro relacionadas à memória, antes de se espalhar para outras regiões (ASSOCIATION, 2020).

A progressão das alterações cerebrais, imperceptíveis para o indivíduo afetado, que causam problemas de memória e, eventualmente, a deficiência física é chamada de DA *continuum* (ilustrado na Figura 1), e inclui três fases distintas, a pré-clínica, o CCL e DA. No estágio pré-clínico, o cérebro apresenta alterações cerebrais mensuráveis que indicam os primeiros sinais da DA (biomarcadores — como níveis anormais de beta amiloide), mas os pacientes ainda não apresentam sintomas identificáveis, como a perda de memória. No próximo estágio, o CCL, além das evidências dos biomarcadores, os pacientes apresentam problemas sutis de memória e de raciocínio, que podem ser percebidos pelos familiares e amigos, mas não por outros, e não interferem na capacidade dos indivíduos de realizar atividades diárias. Na última fase, quando a doença já foi desenvolvida, existem três estágios: o leve, o moderado e o grave, que refletem o grau em que os sintomas interferem na capacidade de realizar atividades cotidianas. Embora o CCL seja considerado um estágio de transição entre o envelhecimento normal e uma síndrome demencial, é importante mencionar que nem todos os pacientes com CCL desenvolvem a DA ou outras demências, sendo a identificação desses casos o foco principal das pesquisas atuais (CUINGNET et al., 2011; ASSOCIATION, 2020; JOHNSON et al., 2012; LUO; TANG, 2017).

Embora existam medicamentos capazes de retardar temporariamente o agravamento dos sintomas da demência e melhorar a qualidade de vida dos pacientes com DA, até o momento ainda não há tratamento disponível para diminuir ou recuperar o dano cerebral causado pela doença (ASSOCIATION, 2020). Consequentemente, as degenerações cerebrais causadas pela DA podem levar o paciente

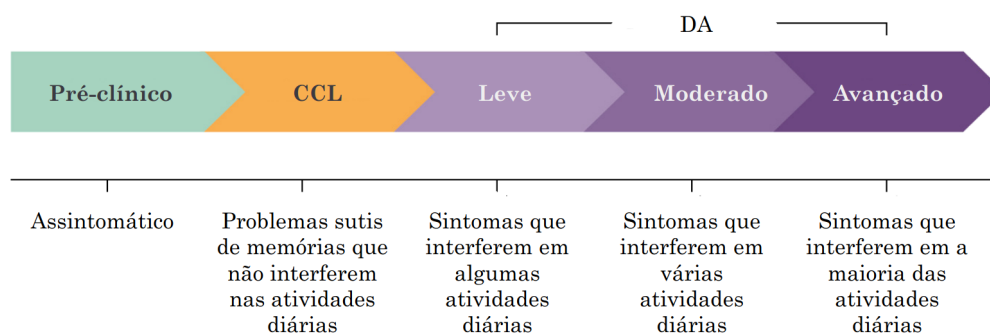


Figura 1 – DA *continuum*. Extraído, editado e traduzido de Association (2020).

ao óbito (JAMES et al., 2014). Estudos indicam que pessoas com 65 anos ou mais sobrevivem em média de 4 a 8 anos após o diagnóstico do Alzheimer, mas algumas vivem até 20 anos. Acredita-se que futuros tratamentos para retardar ou interromper a progressão da DA e preservar a função cerebral possam ser mais eficazes quando administrados no início das alterações cerebrais da doença, ou seja, durante a fase pré-clínica e o CCL (ASSOCIATION, 2020). Além disso, evidências sugerem que, quando as pessoas com demência e suas famílias estão bem preparadas e apoiadas, os sentimentos iniciais de choque, raiva e tristeza são equilibrados por um sentimento de segurança e empoderamento (ALZHEIMER'S DISEASE INTERNATIONAL, 2020).

Até o momento, não existe um único teste diagnóstico definitivo que permita determinar se uma pessoa tem a DA. Portanto, em geral, os médicos (neurologistas, neuropsicólogos, geriatras e psiquiatras geriátricos) usam uma variedade de abordagens e ferramentas para ajudar a realizar o diagnóstico, que incluem a análise de exames clínicos, neuro e psicológicos, imageamento por RM, testes laboratoriais e histórico clínico do paciente (ASSOCIATION, 2020). Os testes neuro e psicológicos podem ajudar a obter sinais de distúrbios da memória, e incluem uma entrevista clínica com o paciente e alguém próximo dele que possa corroborar a perda de memória. Um dos testes mais utilizados se chama *Mini-Mental State Examination* (MMSE), cuja pontuação varia entre 1 e 30 pontos, sendo que quanto mais baixo for o valor do MMSE, maior será a gravidade dos problemas cognitivos do paciente. Os testes laboratoriais, como exames de tireoide e vitamina B_{12} , visam identificar as causas secundárias da demência e distúrbios comuns em idosos

(SCHELTENS et al., 2006), enquanto o imageamento por RM, em conjunto com o histórico clínico, são usados para medir o grau de atrofia no lobo temporal médio (SCHELTENS et al., 1992; DUARA et al., 2008), com análises de alterações volumétricas das estruturas (SCHELTENS et al., 2006; CUINGNET et al., 2011).

A Figura 2 ilustra as principais diferenças encontradas em um cérebro com DA (estágio avançado) e um cérebro saudável. A atrofia cerebral faz com que ocorra um encolhimento de córtex, danificando principalmente regiões envolvidas no raciocínio e lembranças. Esse encolhimento afeta mais gravemente os hipocampos (circulados em branco), regiões responsáveis pelo armazenamento de novas lembranças, e causa um aumento nos ventrículos (circulados em preto).

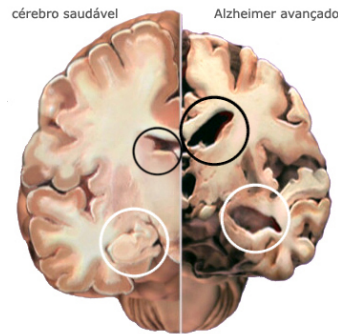


Figura 2 – Diferenças entre um cérebro saudável e com a doença de Alzheimer. Extraído de (ASSOCIATION, 2020).

2.2 Ressonância Magnética

A RM é uma técnica não invasiva e não ionizante que produz imagens estruturais e funcionais. As imagens estruturais são utilizadas rotineiramente na prática clínica, devido à sua capacidade de diferenciar tecidos e fornecer informações sobre a forma e estrutura dos órgãos.

De maneira resumida, uma imagem de RM é produzida a partir do fenômeno de ressonância, que funciona como descrito a seguir. Primeiro, os átomos de hidrogênio do tecido humano são submetidos a um forte campo magnético (provindo do *scanner*) e se orientam na direção do campo (paralela ou antiparalelamente), mantendo o equilíbrio. Em seguida, é inserido um segundo campo, posicionado

perpendicularmente ao primeiro e com oscilação provinda do envio de um impulso de radiofrequência (RF), que faz com que os átomos de hidrogênio deixem de se alinhar ao primeiro campo, criando assim o fenômeno da ressonância. Depois disso, o segundo campo é removido (deixando de atuar no sistema) e os átomos voltam a se alinhar com o primeiro, porém, ao retornar para as suas posições, são geradas ondas magnéticas que variam conforme a intensidade do campo ao qual estão sujeitas. Essas ondas são captadas e analisadas e a partir delas são obtidas informações que correspondem à vizinhança de cada átomo. Por fim, essas informações são utilizadas para a formação da imagem final (BLOCH, 1946; HANSON, 2009; HAGE; IWASAKI, 2009; MAZZOLA, 2009).

As imagens estruturais possuem ainda diferentes ponderações, que traduzem o comportamento imagiológico das estruturas anatômicas e das patologias que podem estar presentes no órgão sendo analisado. Tais ponderações podem ser obtidas pelo controle de determinados parâmetros do *scanner*, dependendo da sequência com que se está trabalhando. As diferentes imagens correspondem à informação que determina a magnitude do sinal adquirido pelo *scanner* e, conseqüentemente, a intensidade de nível de cinza atribuída a cada ponto da imagem.

O valor de intensidade de cinza de cada ponto da imagem é representado por um *voxel* que varia conforme a densidade de prótons de cada tecido — quanto maior a densidade, mais intensa a resposta do sinal de radiofrequência daquela região (HANSON, 2009). O contraste das imagens depende de dois outros parâmetros específicos: T1 e T2, constantes de tempo criadas durante os processos de relaxação durante a formação da imagem. As imagens ponderadas em T1 possuem um melhor delineamento da anatomia do paciente, sendo essa a ponderação mais utilizada para o auxiliar no diagnóstico da DA; e as imagens em T2 fornecem maior contraste às lesões cerebrais e processos inflamatórios presentes, sendo mais utilizadas no auxílio ao diagnóstico de doenças como gliomas e esclerose múltipla.

2.3 Algoritmos de pré-processamento

As subseções a seguir descrevem um conjunto de técnicas de pré-processamento comumente utilizadas no processamento de imagens médicas de RM.

2.3.1 Redução de ruído

As imagens de RM são predominantemente contaminadas por um ruído gerado de flutuações térmicas, que ocorrem devido ao próprio corpo do paciente posicionado no interior do *scanner* e podem interferir negativamente nos demais estágios do processamento. Esse ruído é descrito por duas distribuições gaussianas, sendo uma associada à parte real e outra à parte imaginária dos dados coletados no domínio de Fourier. Após a transformação dos dados para o domínio da imagem, a distribuição do ruído se torna uma função densidade de probabilidade Rician (COUPÉ et al., 2008). A distribuição Rician (GUDBJARTSSON; PATZ, 1995) é uma distribuição contínua para valores positivos (SIJBERS et al., 1998), e que se aproxima de uma distribuição Gaussiana se a relação sinal ruído (SNR) for elevada (FERREIRA, 2013). Conforme relatado em trabalhos (COUPÉ et al., 2008; WOOD; JOHNSON, 1999), o ruído Riciano das imagens de RM pode ser aproximado por um ruído Gaussiano em áreas de alta intensidade, tipicamente encontrada nos tecidos cerebrais. A técnica *Nonlocal Means* (NLM) tem sido bastante utilizada para a atenuação de ruído nas imagens de RM. Ela foi inicialmente proposta por Buades, Coll e Morel (2005) para a redução de ruído aditivo Gaussiano e posteriormente otimizada por Coupé et al. (2008) para a utilização em imagens de RM. A sua implementação está disponível na ferramenta EZMINC¹.

A técnica NLM baseia-se no fato de que uma imagem possui características que se repetem global e localmente, portanto, o método estima a intensidade de cada *voxel* usando uma soma ponderada dos *voxels* da imagem que possuem uma vizinhança (*patch*) similar à do *voxel* analisado. Dessa maneira, o algoritmo consegue reduzir o ruído nas imagens sem incorrer na perda de informações estruturais, conseguindo suavizar pequenas variações nos níveis de cinza da imagem de maneira adaptativa e focando a suavização em regiões que não possuem bordas.

Para calcular a similaridade entre vizinhanças, o algoritmo usa o parâmetro h , que atua como um parâmetro de suavização. Se h for muito alto, todos os *voxels* da vizinhança terão o mesmo peso e haverá uma suavização forte da imagem, ao passo que, se h for muito baixo, apenas alguns *voxels* muito semelhantes terão um peso significativo e haverá uma suavização fraca da imagem. Para

¹ <<https://github.com/BIC-MNI/EZmnc>>

estimar o valor de h deve-se conhecer o desvio-padrão do ruído presente na imagem, o qual, em Coupé et al. (2008) foi estimado a partir do cálculo do pseudo-resíduo (GASSER; SROKA; JENNEN-STEINMETZ, 1986; BOULANGER; KERVANN; BOUTHEMY, 2005; COUPÉ et al., 2008). Além do ajuste automático do parâmetro de suavização, os autores também implementaram uma seleção dos *voxels* mais relevantes e uma implementação em blocos que usa programação paralela.

2.3.2 Correção de *bias field*

Além do ruído Riciano, as imagens de RM, em especial as produzidas por *scanners* antigos de RM, também são corrompidas por um sinal de baixa frequência chamado de *bias field*. Essa degradação é um fenômeno adverso que pode ser causado por uma série de fatores, como heterogeneidades na sensibilidade das bobinas de radiofrequência, variações no campo magnético estático, entre outros.

O algoritmo N4-ITK (TUSTISON et al., 2010), disponibilizado pela biblioteca ITK, tem sido utilizado para a correção do *bias field*. O N4 é uma variante do algoritmo *Nonparametric Nonuniform intensity Normalization* (N3) (SLED; ZIJDENBOS; EVANS, 1998). Em ambos os métodos, N3 e N4-ITK, o modelo matemático de formação da imagem de RM assume que o *bias field* é um sinal multiplicativo que suaviza o sinal originalmente emitido pelo tecido do paciente em uma determinada localidade. O N3 é um algoritmo iterativo que busca por um campo multiplicativo que maximize o conteúdo de alta frequência da distribuição de intensidade do tecido. Já o N4, utiliza uma otimização hierárquica modificada e um algoritmo de aproximação *B-Spline* mais rápido e robusto.

2.3.3 Padronização de histogramas

O uso de diferentes bases de imagens e equipamentos de RM podem resultar em uma distribuição não padronizada dos níveis de cinza para um mesmo tecido nas imagens T1- w de pacientes distintos. Essa falta de padronização na distribuição de intensidade é um dos principais problemas do processamento de imagens de RM e dificulta na aplicação de algoritmos baseados na intensidade dos *voxels* (NYÚL; UDUPA; ZHANG, 2000). Para a correção das distribuições de intensidade, foi

utilizada uma variação do algoritmo proposto por Nyúl, Udupa e Zhang (2000), disponível na biblioteca ITK. O algoritmo usa uma imagem de referência para corrigir a distribuição de intensidade de novas imagens, utilizando uma técnica similar à técnica convencional de casamento de histogramas. Assim, as distribuições dos níveis de cinza de todas as imagens se tornam similares a da imagem de referência.

2.3.4 Corregistro das imagens

Para garantir um processamento correto nas demais etapas, todas as imagens devem estar alinhadas a um mesmo sistema de coordenadas (origem, orientação) e possuir as mesmas resoluções. Para isso, são utilizadas técnicas de corregistro afim e deformável. O corregistro afim proposto por (OURSELIN et al., 2001) e (OURSELIN; STEFANESCU; PENNEC, 2002) e disponível na ferramenta Nifty Reg², tem sido bastante utilizado por apresentar bons resultados em diferentes situações. Esse corregistro, que envolve apenas transformações globais, como rotação, translação, escala e cisalhamento, utiliza uma abordagem iterativa baseada no casamento de blocos, que busca por um conjunto de blocos correspondentes entre uma imagem de referência e uma imagem móvel. O corregistro deformável, disponibilizado pela mesma biblioteca, é baseado em transformações *B-Spline* e uma abordagem multirresolução baseado na deformação *Free-From* apresentada por Rueckert et al. (1999) e posteriormente otimizada para o modelo de Modat et al. (2010).

2.3.5 Extração da máscara do encéfalo

Para delimitar a região de análise no encéfalo, foi utilizada a técnica *Robust Brain Extraction* (ROBEX) (IGLESIAS et al., 2011), que combina um modelo discriminativo e um modelo geral para a extração da região. No modelo discriminativo é utilizado um classificador *Random Forest* já treinado para detectar as bordas do encéfalo. O modelo geral, utilizado quando uma nova imagem é apresentada ao ROBEX, buscará pelo contorno que possui a maior probabilidade de ser o contorno do encéfalo daquela imagem, conforme o modelo discriminativo

² <http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg>

já treinado. Um ponto importante sobre o ROBEX é que ele foi projetado para funcionar sem a necessidade de ajuste de parâmetros.

2.4 Filtros Log-Gabor 3D

Os filtros log-Gabor 3D seguem a formulação matemática proposta por Dosil, Pardo e Fdez-Vidal (2005). Além de apresentar componentes DC (*direct current*) nulas, portanto, não responder a regiões com intensidades constantes, eles se estendem até as altas frequências, o que permite obter uma ampla largura de banda que pode fornecer uma cobertura mais uniforme do domínio da frequência.

Cada filtro log-Gabor 3D é definido como o resultado da multiplicação de um termo radial e um termo angular, sendo

$$G(\boldsymbol{\rho}; \omega_s, \eta_\beta, \theta_i, \phi_{i,j}) = \underbrace{\exp\left(-\frac{\log^2(\|\boldsymbol{\rho}\|/\omega_s)}{2\log^2\eta_\beta}\right)}_{\text{radial}} + \underbrace{\exp\left(-\frac{\alpha(\boldsymbol{\rho}, \theta_i, \phi_{i,j})}{2\sigma_\alpha^2}\right)}_{\text{angular}}, \quad (1)$$

em que $\alpha(\boldsymbol{\rho}, \theta_i, \phi_{i,j}) = \arccos^2\left(\frac{\mathbf{v} \cdot \boldsymbol{\rho}}{\|\boldsymbol{\rho}\|}\right)$, o parâmetro $\boldsymbol{\rho} = (u, v, w)$ é um ponto (vetor) no domínio da frequência, e $\mathbf{v} = (\cos\phi_0 \cos\theta_0, \cos\phi_0 \sin\theta_0, \sin\theta_0)$ contém os ângulos do sistema de coordenadas esféricas, com θ_i e $\phi_{i,j}$ representando os ângulos de elevação e azimute, respectivamente. A combinação dos ângulos de elevação e azimute formam uma orientação Θ . Adicionalmente, a cobertura de frequência do filtro pode ser modificada com a alteração do parâmetro ω_s , o qual representa a frequência central para cada escala $s = \{0, 1, \dots, N_s - 1\}$, $\eta_\beta = \sigma_\omega/\omega_s$ e σ_α , que controlam a largura de banda e o espalhamento angular do filtro, respectivamente.

A frequência central ω_s é calculada como $\omega_s = \omega_{\max}/\gamma^s$, em que γ é um fator de escala (na direção radial) entre sucessivos filtros. O ângulo de elevação é calculado como

$$\theta_i = \frac{\pi i}{2(N_e - 1)}, \quad (2)$$

em que $i = \{0, 1, \dots, N_e - 1\}$ e N_e correspondem respectivamente aos índices e número de ângulos de elevação. Para a implementação do banco de filtros log-Gabor, é utilizado um valor de comprimento de arco igual entre os valores de ângulo azimute adjacentes sobre uma esfera de raio unitário, ao invés de tomar distâncias angulares uniformes (DOSIL; PARDO; FDEZ-VIDAL, 2005). Dessa

maneira, o ângulo de elevação é amostrado uniformemente e o valor do ângulo azimute varia em função da elevação, mantendo a densidade dos filtros constante, sendo

$$N_a(i) = N_{a,0} \cos \theta_i, \quad (3)$$

em que $N_{a,0}$ é o número de ângulos de azimute no equador de uma esfera de raio unitário. Assim que o número de azimutes por elevação é obtido, o azimute central $\phi_{i,j}$, para a i^{a} elevação, é calculado como

$$\phi_{i,j} = \begin{cases} \frac{\pi j}{N_{a,0}}, & \text{se } i = 0 \\ \frac{2\pi j}{N_a(i)}, & \text{senão} \end{cases}, \quad (4)$$

em que $j = \{0, 1, \dots, N_a(i) - 1\}$.

2.5 Momentos de segunda ordem

O mapa de momentos locais é calculado com base na teoria de análise de momentos discretos de imagens (LO; DON, 1989), em que é calculado um mapa de momentos tridimensionais da ordem $p + q + r$ conforme a fórmula

$$M_{p,q,r}(\mathbf{x}) = \sum_{\Theta} x_{\Theta}^p(\mathbf{x}) y_{\Theta}^q(\mathbf{x}) z_{\Theta}^r(\mathbf{x}), \quad (5)$$

em que

$$\begin{aligned} x_{\Theta}(\mathbf{x}) &= IM_{\Theta}(\mathbf{x}) \cos \theta_i \cos \phi_{i,j}, \\ y_{\Theta}(\mathbf{x}) &= IM_{\Theta}(\mathbf{x}) \cos \theta_i \sin \phi_{i,j}, \\ z_{\Theta}(\mathbf{x}) &= IM_{\Theta}(\mathbf{x}) \sin \theta_i. \end{aligned} \quad (6)$$

e IM_{Θ} corresponde a uma imagem direcional de orientação Θ e $\mathbf{x} = (i, j, k)$ à localização espacial no domínio da imagem. A matriz de momentos de segunda ordem de inércia (Equação 5) fornece uma noção de covariância dos momentos de IM nas variáveis $x_{\Theta}(\mathbf{x})$, $y_{\Theta}(\mathbf{x})$ e $z_{\Theta}(\mathbf{x})$ e pode ser escrita como

$$M_{IM}(\mathbf{x}) = \begin{bmatrix} M_{2,0,0}(\mathbf{x}) & M_{1,1,0}(\mathbf{x}) & M_{1,0,1}(\mathbf{x}) \\ M_{1,1,0}(\mathbf{x}) & M_{0,2,0}(\mathbf{x}) & M_{0,1,1}(\mathbf{x}) \\ M_{1,0,1}(\mathbf{x}) & M_{0,1,1}(\mathbf{x}) & M_{0,0,2}(\mathbf{x}) \end{bmatrix}. \quad (7)$$

O principal momento de inércia pode ser aproximado pelos autovalores da matriz $M_{IM}(\mathbf{x})$ (sendo denotados por $\lambda_0(\mathbf{x}) \geq \lambda_1(\mathbf{x}) \geq \lambda_2(\mathbf{x}) \geq 0$) e fornecem um senso de direção local (na posição \mathbf{x}) ao longo da qual os dados têm a maior variação.

2.6 Filtragem direcional

As imagens direcionais são calculadas a partir da convolução de uma imagem com os filtros log-Gabor 3D para cada orientação individual Θ e escala s . Em seguida, as escalas são combinadas, para cada orientação, com o objetivo de formar uma imagem direcional resultante $DR_{\Theta}(\mathbf{x})$, da seguinte maneira

$$DR_{\Theta}(\mathbf{x}) = \max_{0 \leq s < N_s} (A_{\Theta_s}(\mathbf{x})), \quad (8)$$

em que $A_{\Theta_s}(\mathbf{x})$ corresponde a magnitude entre os *voxels* de todas as escalas, e $\mathbf{x} = (i, j, k)$ corresponde a localização espacial no domínio da imagem.

A partir das imagens resultantes, $DR_{\Theta}(\mathbf{x})$, foram calculados os momentos locais de segunda ordem da matriz $M_{DR}(\mathbf{x})$. Nesse caso, a IM_{Θ} da matriz $M_{IM}(\mathbf{x})$ da Equação 6 foi substituída por $DR_{\Theta}(\mathbf{x})$. Como os autovalores da matriz $M_{DR}(\mathbf{x})$ fornecem um senso de direção local, foram armazenadas as imagens que contém as respostas dos filtros direcionais DR_{Θ} , em que $\Theta = \{\Theta_0, \Theta_1, \Theta_2 \dots, \Theta_{22}\}$ e os mapas de autovalores $\lambda_0(\mathbf{x})$, $\lambda_1(\mathbf{x})$, e $\lambda_2(\mathbf{x})$.

A Figura 3 mostra a vista sagital do cérebro como uma representação visual da etapa de filtragem direcional. Os quatro valores de Θ ilustrados na imagem correspondem ao primeiro ângulo de azimute combinado com cada uma das quatro elevações (por exemplo, $\Theta_0 = (\phi_0, \theta_0)$, ..., $\Theta_{22} = (\phi_0, \theta_3)$).

2.7 Congruência de fase 3D

A técnica de Congruência de Fase (*Phase Congruency* — PC) 3D, proposta por Villa-Pinto e Ferrari (2016), é uma extensão para 3D da PC proposta por Kovess (2000) que, por sua vez, é baseada no modelo de energia proposto por Morrone et al. (1986). Diferente de outras técnicas para a detecção de pontos salientes, como a técnica SIFT (LOWE, 2004; RUECKERT et al., 1999) e o detector de Harris e

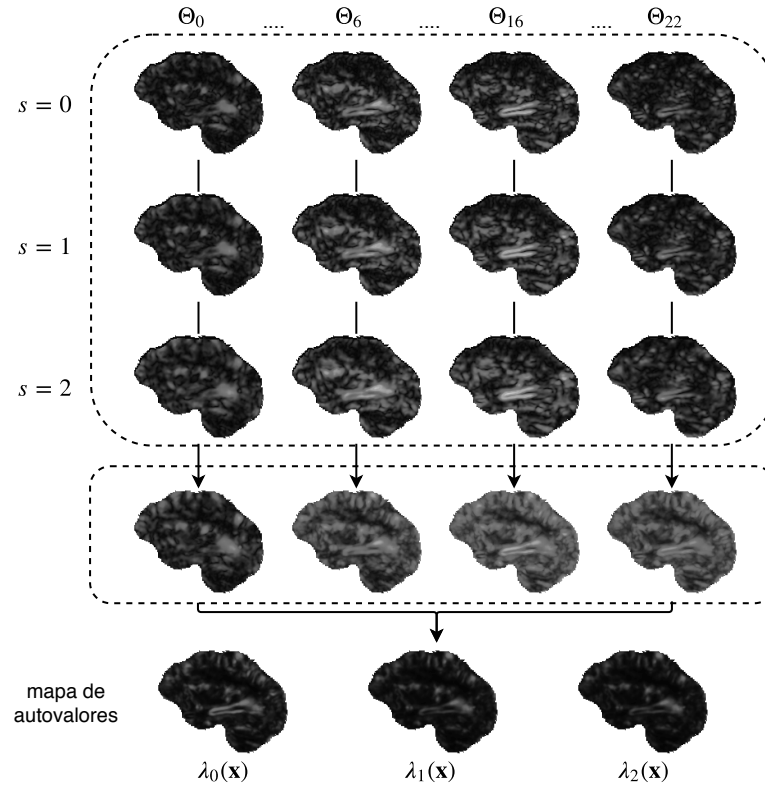


Figura 3 – Representação visual das respostas dos filtros na vista sagital.

Stephens (1988), que detectam pontos salientes como posições na imagem com alta variação de intensidade (gradiente), a técnica de PC detecta os pontos salientes como posições nas quais os componentes de Fourier se encontram maximalmente em fase. Além disso, a medida PC é baseada principalmente na informação de fase do sinal, provendo robustez em relação às variações de intensidade da imagem.

A medida PC é comprovadamente proporcional à energia local de um sinal e desse modo, pode ser calculada pela convolução da imagem original com um banco de filtros espaciais em quadratura (VENKATESH; OWENS, 1990), ou seja, filtros que possuem um desvio de fase de 90° , como o banco de filtros log-Gabor 3D. Sendo assim, a congruência de fase é calculada após a convolução de uma imagem com os filtros log-Gabor 3D para cada orientação individual Θ .

A técnica de PC pode ser estendida para uso em 3D combinando a técnica 1D em um conjunto de orientações $\Theta = (\phi, \theta)$, em que ϕ e θ correspondem ao ângulo

azimute e elevação, respectivamente

$$PC(\mathbf{x}) = \frac{\sum_{\Theta} \sum_n W_{\Theta}(\mathbf{x}) \max(0, A_{\Theta n}(\mathbf{x}) \Delta \Phi_{\Theta n}(\mathbf{x}) - T)}{\sum_{\Theta} \sum_n A_{\Theta n}(\mathbf{x}) + \varepsilon}, \quad (9)$$

em que $\mathbf{x} = (i, j, k)$ corresponde à localização espacial no domínio da imagem, a constante T é um limiar estimado automaticamente que representa a energia do ruído, e ε a uma constante de valor pequeno utilizada para evitar a divisão por zero (VILLA-PINTO; FERRARI, 2016). O termo $A_{\Theta n}(\mathbf{x})$, definido como

$$A_{\Theta n}(\mathbf{x}) = \sqrt{e_{\Theta n}(\mathbf{x})^2 + o_{\Theta n}(\mathbf{x})^2}, \quad (10)$$

corresponde a amplitude da energia e $e_{\Theta n}(\mathbf{x}) = I * G_{\Theta n}^e$ e $o_{\Theta n}(\mathbf{x}) = I * G_{\Theta n}^o$ são obtidos via convolução da imagem de entrada (I) com um par de filtros em quadratura ($G_{\Theta n}^e, G_{\Theta n}^o$), com escala n e orientação indicada pelo vetor Θ . O termo $W_{\Theta}(\mathbf{x})$ é uma função sigmoide que penaliza distribuições de frequências estreitas, definido como

$$W_{\Theta}(\mathbf{x}) = \frac{1}{1 + \exp\{\gamma [c - l_{\Theta}(\mathbf{x})]\}}, \quad (11)$$

em que os parâmetros γ e c são constantes que representam fatores de ganho e corte, respectivamente. O termo $l_{\Theta}(\mathbf{x})$ é uma métrica utilizada para a obtenção da largura da distribuição de resposta do filtro

$$l_{\Theta}(\mathbf{x}) = \frac{1}{N_s} \left(\frac{\sum_n A_{\Theta n}(\mathbf{x})}{A_{\Theta \max}(\mathbf{x}) + \varepsilon} \right), \quad (12)$$

em que $A_{\Theta \max}(\mathbf{x})$ corresponde a maior resposta obtida ao longo da variação das escalas. O termo $\Delta \Phi_{\Theta n}(\mathbf{x})$, da Equação 9, é a função de desvio de fase utilizada para aumentar a “nitidez” das respostas PCs, definido como

$$\begin{aligned} \Delta \Phi_{\Theta n}(\mathbf{x}) &= \frac{\cos(\phi_{\Theta n}(\mathbf{x}) - \bar{\phi}_{\Theta}(\mathbf{x})) - |\sin(\phi_{\Theta n}(\mathbf{x}) - \bar{\phi}_{\Theta}(\mathbf{x}))|}{e_{\Theta n}(\mathbf{x}) \left(\frac{\sum_n e_{\Theta n}(\mathbf{x})}{E_{\Theta}(\mathbf{x})} \right) + o_{\Theta n}(\mathbf{x}) \left(\frac{\sum_n o_{\Theta n}(\mathbf{x})}{E_{\Theta}(\mathbf{x})} \right) - |e_{\Theta n}(\mathbf{x}) \left(\frac{\sum_n o_{\Theta n}(\mathbf{x})}{E_{\Theta}(\mathbf{x})} \right) - o_{\Theta n}(\mathbf{x}) \left(\frac{\sum_n e_{\Theta n}(\mathbf{x})}{E_{\Theta}(\mathbf{x})} \right)|}, \end{aligned} \quad (13)$$

em que

$$E_{\Theta}(\mathbf{x}) = \sqrt{\left(\sum_n e_{\Theta n}(\mathbf{x}) \right)^2 + \left(\sum_n o_{\Theta n}(\mathbf{x}) \right)^2}, \quad (14)$$

corresponde a energia local da imagem ao longo de todas as escalas.

2.7.1 Detecção de pontos salientes

A detecção dos pontos salientes utiliza os momentos locais de segunda ordem, calculados a partir dos mapas de momentos de segunda ordem de inércia, obtidos a partir de cada mapa direcional da PC. Nesse caso, a matriz de momentos $M_{IM}(\mathbf{x})$ da equação 6 foi substituída pela $M_{PC}(\mathbf{x})$, visto que a IM_{Θ} corresponde aos mapas de PC_{Θ} . Como os principais momentos de inércia possuem uma relação direta com os autovalores da matriz $M_{PC}(\mathbf{x})$ ($\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq 0$), Villa-Pinto e Ferrari (2016) propuseram utilizar o valor normalizado do menor autovalor, λ_2 , calculado por

$$\text{Saliência}(\mathbf{x}) = \lambda_2 / \lambda_{2\max}, \quad (15)$$

como uma métrica para determinar os pontos salientes. Dessa maneira, para cada localização \mathbf{x} na imagem, os autovalores de $M_{PC}(\mathbf{x})$ são calculados e o *voxel* em \mathbf{x} é considerado um ponto saliente se a Saliência(\mathbf{x}) for maior que um limiar T_{λ} .

2.7.2 Descritores

Com o objetivo de encontrar casamentos entre pares de pontos detectados, Villa-Pinto e Ferrari (2016) propuseram a análise de similaridade dos pontos salientes de diferentes imagens usando um descritor (assinatura) de região, o qual é baseado na abordagem *shape context* (BELONGIE; MALIK; PUZICHA, 2002; MORTENSEN; DENG; SHAPIRO, 2005) e busca quantificar informações de forma. A proposta original do descritor *shape context*, para imagens 2D, tem o objetivo de quantificar semelhanças de forma e encontrar correspondências locais entre objetos que possuem seus contornos representados por conjuntos de pontos 2D.

A adaptação da técnica para 3D proposta por Villa-Pinto e Ferrari (2016) envolve a utilização do mapa dos maiores autovalores (λ_0) da matriz $M_{PC}(\mathbf{x})$ como imagem fonte para o cômputo do descritor, dado que os maiores autovalores conferem uma noção de localização e significância da informação de borda contida na imagem original. As abordagens global e local do descritor *shape context* são descritas a seguir.

2.7.2.1 *Shape context* 3D global

Villa-Pinto e Ferrari (2016) criaram um histograma log-esférico limitado por uma distância radial máxima, r_{max} , que divide o espaço da imagem ao redor de cada ponto em η_r distâncias radiais (e logarítmicas) relativas, η_a ângulos de azimute e η_e ângulos de elevação. As divisões radiais do histograma são definidas pelos valores $r_{max}/2^{\eta_r-j}$, para $j = \eta_r - 1, \eta_r - 2, \dots, 1$.

O histograma log-esférico, \mathbf{h} , acumula valores que expressam a significância dos pontos detectados para cada setor e a concatenação desses setores forma o descritor final. Devido à natureza da divisão de r_{max} , esse descritor consegue captar características globais, pois setores mais próximos do ponto saliente possuem menor volume, e conseqüentemente acumulam menos informação de borda, do que setores mais distantes.

2.7.2.2 *Shape context* 3D local

Baseado na metodologia proposta por (ANKERST et al., 1999) e utilizando o mesmo mapa de bordas que o descritor *shape context* 3-D global, foi criado um segundo descritor, com a divisão do espaço esférico seguindo um modelo *spider-web*. De maneira similar ao *shape context* 3D global, também foi criado um histograma log-esférico, no entanto, dessa vez as divisões radiais do histograma são limitadas pelos valores $r_{max}/2^{\eta_r-j}$, para $j = 1, \dots, \eta_r - 2, \eta_r - 1$.

Devido ao modelo *spider-web*, obtém-se um descritor capaz de captar características locais, visto que setores mais próximos do ponto saliente possuem maior volume, e conseqüentemente, mais informação de borda do que os mais distantes.

A Figura 4 ilustra a representação gráfica do espaço log-esférico nos dois descritores *shape context* 3D. O global é representado pela Figura 4(a) e o local pela 4(b).

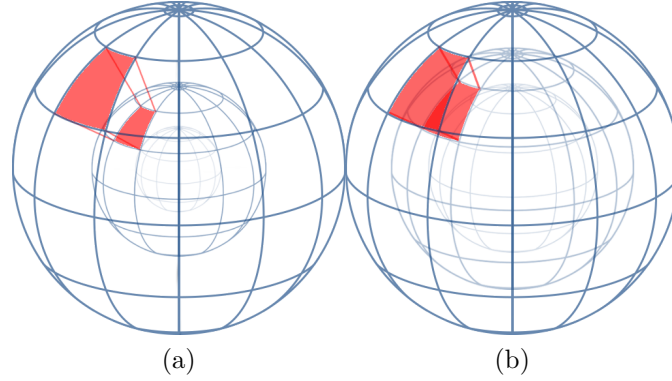


Figura 4 – Representação gráfica do espaço log-esférico ao redor de um ponto saliente para o cômputo do descritor. (a) *shape context* 3D global; (b) *shape context* 3D local.

2.7.3 Casamento

Para medir a semelhança entre pares de descritores foi utilizada a distância χ^2 como função de similaridade, calculada como

$$\chi^2(\mathbf{h}_1, \mathbf{h}_2) = \frac{1}{2} \sum_{b=0}^{\eta_r \eta_a \eta_e - 1} \frac{(h_1^{(b)} - h_2^{(b)})^2}{h_1^{(b)} + h_2^{(b)}}, \quad (16)$$

em que $h_1^{(b)}$ e $h_2^{(b)}$ correspondem, respectivamente, ao b -ésimo componente de \mathbf{h}_1 e \mathbf{h}_2 . A distância χ^2 é uma medida estatística considera as diferenças entre os tamanhos das classes dos histogramas, fazendo com que pequenas diferenças entre os valores obtidos por classes maiores (que acumulam valores muito maiores) resultem em uma distância menor, e vice-versa (BELONGIE; MALIK; PUZICHA, 2002).

Após computar as distâncias de descritores entre todos os pares de pontos salientes das imagens de entrada, os pontos correspondentes são encontrados e casados com base em uma busca pelo vizinho mais próximo nos sentidos “ida e volta”. Neste caso, assume-se que \mathbf{h}_1 e \mathbf{h}_2 são casados se \mathbf{h}_2 é o vizinho mais próximo de \mathbf{h}_1 e \mathbf{h}_1 é também o vizinho mais próximo de \mathbf{h}_2 (VILLA-PINTO; FERRARI, 2016).

2.7.4 Detecção do plano sagital médio

A técnica de detecção do plano sagital médio (MSP), proposta por Ferrari, Villa-Pinto e Moreira (2016), determina os *voxels* com maior probabilidade de pertencer ao MSP a partir de uma medida de *sheetness*, obtida da análise dos autovalores da matriz de segunda ordem da congruência da fase 3D. Utilizando os *voxels* com os maiores valores de *sheetness*, determinados via um limiar fixo, um algoritmo de ajuste de mínimos quadrados ponderados (*Weighted Least Squares* — WLS) (MONTGOMERY; PECK; VINING, 2012) é usado para determinar o plano inicial. Esse plano é então refinado iterativamente a medida que os pontos mais distantes do plano estimado na iteração anterior são eliminados, resultando em um novo conjunto de pontos para a estimação na próxima iteração. Nesse caso, os valores da medida de *sheetness* são usados como pesos pelo algoritmo de ajuste do WLS. O procedimento iterativo trabalha com margem decrescente de 2 *voxels* e termina quando a margem é reduzida para 3 *voxels*.

2.8 Atlas probabilístico de pontos salientes 3D

O atlas probabilístico de pontos salientes 3D tem como objetivo fornecer pontos salientes mais estáveis e descritivos de uma população. Para a sua criação, foi utilizada a metodologia proposta por Poloni et al. (2018).

A construção do atlas probabilístico de pontos salientes 3D é realizada a partir de um conjunto de imagens de pacientes saudáveis com seus pontos salientes 3D e descritores computados. Sua criação é realizada a partir da estimativa de uma função densidade de probabilidade, $p(\mathbf{x}) : X \subset \mathbb{N}^3 \rightarrow [0, 1]$, que representa a densidade espacial de todos os pontos detectados nas imagens de treinamento. Essa estimativa é realizada utilizando o modelo não paramétrico *Manifold Parzen Windows* (MPW) (VINCENT; BENGIO, 2003), em que $p(\mathbf{x})$ representa um conjunto de amostras como uma distribuição de misturas de Gaussianas, $G_{\mu, \Sigma}$, e expressa

a ocorrência dos pontos na população como

$$p(\mathbf{x}) = \frac{1}{l} \sum_{i=1}^l G_{\mu=\mathbf{x}_i, \Sigma=\Sigma_i}(\mathbf{x}) \quad (17)$$

$$= \frac{1}{l} \sum_{i=1}^l \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_i)}} \exp \left[-\frac{1}{2} \Delta(\mathbf{x}) \right], \quad (18)$$

em que l representa o número de amostras, d a dimensão, $\Delta(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{x}_i)$ é a distância quadrada de *Mahalanobis* e $\Sigma_i = \Sigma_{\mathcal{K}_i} + \sigma^2 \mathbf{I}_3$ representa a matriz de covariância do i -ésimo componente Gaussiano. Como a inversa da matriz de covariância, Σ_i^{-1} , pode se tornar um problema mal condicionado, $\sigma^2 \mathbf{I}_3$ foi adicionada ao termo $\Sigma_{\mathcal{K}_i}$, em que $\sigma^2 \mathbf{I}_3$ corresponde a uma matriz 3×3 gaussiana isotrópica com $\sigma^2 = 1$. O primeiro componente $\Sigma_{\mathcal{K}_i}$ da matriz de covariância Σ_i é calculado como

$$\Sigma_{\mathcal{K}_i} = \frac{\sum_{j=1 \dots l, j \neq i} \mathcal{K}(\mathbf{x}_j, \mathbf{x}_i) (\mathbf{x}_j - \mathbf{x}_i)^t (\mathbf{x}_j - \mathbf{x}_i)}{\sum_{j=1 \dots l, j \neq i} \mathcal{K}(\mathbf{x}_j, \mathbf{x}_i)}, \quad (19)$$

sendo $\mathcal{K}(\mathbf{x}_j, \mathbf{x}_i)$ uma função de vizinhança que associa a cada amostra \mathbf{x}_j de seu vizinho \mathbf{x}_i um fator de ponderação. Tal vizinhança pode ser limitada por k_{cov} amostras. O cálculo da densidade em cada ponto saliente pode ser simplificado se a combinação de funções gaussianas for restrita a apenas k componentes, que correspondem aos k pontos vizinhos mais próximos.

Após estimar a densidade, para cada *voxel* correspondente a um ponto maximal de $p(\mathbf{x})$, um ponto saliente médio é obtido e seu descritor *shape context* (local ou global) médio $\bar{\mathbf{h}}$ é calculado, sendo esse correspondente à média de todos os descritores computados em todas as imagens dentro de uma região vizinha em torno do pico maximal correspondente (POLONI et al., 2018). Como resultado da criação do atlas, tem-se um mapa (uma imagem) de densidades cujos valores são proporcionais à presença de pontos salientes considerando a população analisada.

2.9 Segmentação de tecidos

A ferramenta *FMRIB's Automated Segmentation Tool* — FAST, desenvolvida por Zhang, Brady e Smith (2001), é considerada o estado da arte em segmentação dos tecidos cerebrais (REHMAN; HWANG; LEE, 2020). Ela segmenta as imagens

de RM em diferentes de tecidos, como GM, substância branca (*White Matter* — WM) e líquido cefalorraquidiano (*Cerebrospinal fluid* — CSF), além de corrigir variações espaciais de intensidades.

O método implementado pela FAST é baseado no modelo *Hidden Markov Random Field* (HMRF) e no algoritmo *Expectation Maximization* (EM) e está disponível no pacote FSL³. O HMRF representa um processo estocástico gerado por um campo aleatório Markoviano, *Markov Random Field* (MRF), cuja sequência de estados não pode ser observada diretamente, mas pode ser estimada indiretamente por observações. A vantagem do modelo HMRF deriva da maneira como a informação espacial é codificada, pois, são consideradas as influências mútuas dos vizinhos locais. Com isso, os *pixels* vizinhos têm os mesmos rótulos de classe (no caso de imagens constantes) ou intensidades semelhantes (no caso de imagens contínuas). O EM foi utilizado para estimar os parâmetros do modelo.

2.10 Classificador SVM

O SVM é um classificador determinístico baseado na teoria da aprendizagem estatística, desenvolvido com o intuito de resolver problemas de aprendizagem de máquinas com pequenas amostras (VAPNIK, 1999). O SVM apresenta vantagens importantes sobre outros classificadores, principalmente em relação a sua habilidade em resolver problemas de classificação com pequeno número de amostras e conseguir boa generalização em espaços altamente dimensionais (ABE, 2005).

O objetivo do SVM é encontrar o hiperplano que melhor separe as amostras de treinamento, geralmente com um mapeamento implícito das amostras de treinamento do espaço de entrada para um espaço de maior dimensão (espaço de características). Idealmente, no novo espaço, as amostras serão linearmente separáveis por um hiperplano, mesmo que não sejam linearmente separáveis no espaço de entrada (WEBB, 2002). Entretanto, podem existir infinitos hiperplanos para a separação das amostras. Sendo assim, o SVM busca, via processo de otimização, pelo hiperplano que possui a maior distância geométrica ou a maior margem de separação entre os dados de duas classes, o qual é denominado hiperplano ótimo

³ <<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki>>

(ZHENG; XUE, 2009).

A formulação do SVM permite que sua margem seja mais (ou menos) flexível, permitindo que amostras localizadas entre as margens sejam penalizadas com menor (margem suave) ou maior (margem rígida) intensidade. A flexibilização da margem é realizada usando as variáveis de folga que permitem a existência de amostras na margem e possivelmente alguns erros de classificação. As amostras localizadas sobre da margem, consideradas as mais informativas dos dados de treinamento, são denominadas de *support vectors* (LORENA; CARVALHO, 2007).

Sendo assim, o SVM maximiza a margem (encontrar o hiperplano ótimo) ao mesmo tempo que penaliza as amostras que se encontram do lado errado do seu limite. Para isso, é introduzido o parâmetro de penalização C , que controla o compromisso entre os erros das variáveis de folga e a generalização do classificador (margem) (BISHOP, 2006). De maneira geral, quanto maior for o valor do parâmetro C , menor será o erro de treinamento final. Porém, se o valor de C for muito alto, o classificador pode perder suas propriedades de generalização, visto que ele tentará ajustar da melhor maneira possível todas as amostras de treinamento, permitindo a inclusão de possíveis erros ou ruídos presentes no seu conjunto de amostras. Além disso, um valor de C grande geralmente aumenta o tempo necessário para treinamento (NELL; SHAW-TAYLOR, 2000).

O mapeamento implícito das amostras de treinamento para um espaço de maior dimensão é realizado com a aplicação de uma função não linear, a qual pode ser representada por uma matriz *kernel*, desde que esta satisfaça determinadas condições (ABE, 2005). A escolha apropriada do mapeamento faz com que o conjunto de dados mapeado possa ser separado por um hiperplano com margem máxima no novo espaço e crie uma borda de decisão não linear no espaço de entrada (LORENA; CARVALHO, 2007; SCHOLKOPF; SMOLA, 2001; ZHENG; XUE, 2009). Alguns dos *kernels* mais utilizados no SVM, que satisfazem essas condições, são o polinomial, a função de base radial (RBF) e o sigmoide.

2.11 Testes estatísticos

Nesta seção são apresentados os testes estatísticos utilizados nesta pesquisa.

2.11.1 Welch *test two-tailed*

O teste Welch *test two-tailed* é utilizado para comparar as médias de dois grupos independentes com a premissa de que os dados das duas populações vêm de uma distribuição normal (NAVARRO, 2019). O teste assume como hipótese nula (H_0) que as duas médias são iguais e como hipótese alternativa (H_A) que a média de um grupo é maior ou menor que a do outro, definidas formalmente como

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_A : \mu_1 &< \mu_2 \mid \mu_2 < \mu_1 \end{aligned} \quad (20)$$

O Welch *test* (WELCH, 1947) é calculado como

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}(\bar{X}_1 - \bar{X}_2)}, \quad (21)$$

em que \bar{X}_1 e \bar{X}_2 correspondem, respectivamente, às médias dos grupos 1 e 2 e SE corresponde ao erro padrão, calculado como

$$\text{SE}(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}, \quad (22)$$

sendo que N_1 e N_2 representam o número de amostras de cada grupo e σ_1^2 e σ_2^2 às suas respectivas variâncias.

Para determinar se o valor obtido para a estatística t rejeita ou aceita H_0 , são calculados o p-valor e o valor crítico da função densidade probabilidade (pdf) t de *Student*, uma distribuição simétrica e similar à distribuição normal, com a distinção de possuir caudas mais largas definidas pelo parâmetro v . O valor de v é definido pelo grau de liberdade df , sendo

$$df = \frac{(\sigma_1^2/N_1 + \sigma_2^2/N_2)}{(\sigma_1^2/N_1)/(N_1 - 1) + (\sigma_2^2/N_2)/(N_2 - 1)}. \quad (23)$$

O valor crítico, que representa um ponto sobre a distribuição estatística, é obtido a partir de um nível de significância (α) e de v . Em seguida é calculado o p-valor, que corresponde à área sobre a curva da pdf t de *Student* para valores maiores que o valor crítico. Por fim, é verificado se H_0 será aceita ou rejeitada, sendo

$$\begin{cases} H_0, & \text{se p-valor} > \alpha \\ H_1, & \text{senão} \end{cases} \quad (24)$$

2.11.2 ANOVA

O teste estatístico utilizado neste trabalho para a comparação de duas ou mais médias é chamado de *One-Way* ANOVA. O teste “*One-Way*” é utilizado quando existe apenas uma variável categórica explicatória dos dados (SELTMAN, 2012) e o ANOVA é utilizado para variáveis independentes, com a premissa de que os resíduos do modelo vêm de uma distribuição normal e suas variâncias são homogêneas (NAVARRO, 2019). O teste assume como hipótese nula (H_0) que todas as médias são iguais, e como hipótese alternativa (H_A) que existe pelo menos uma média que difere, definidas formalmente como

$$\begin{aligned} H_0 : \mu_1 = \mu_2 = \dots = \mu_k \\ H_A : \exists_{i,j} : \mu_i \neq \mu_j \end{aligned} \quad (25)$$

A ANOVA está associada a soma de quadrados total (SS_{tot}), que mede a variabilidade total dos dados, sendo

$$SS_{tot} = \sum_{j=1}^K \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y})^2, \quad (26)$$

em que K refere-se à quantidade de grupos que a variável categórica possui e N ao total de amostras. A equação 26 pode ser dividida em duas partes, a soma de quadrados dentro SS_D , e a soma dos quadrados entre SS_E grupos

$$SS_{tot} = SS_D + SS_E, \quad (27)$$

sendo

$$SS_D = \sum_{j=1}^K \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y}_j)^2, \quad (28)$$

e

$$\begin{aligned} SS_E &= \sum_{j=1}^K \sum_{i=1}^{N_j} (\bar{Y}_j - \bar{Y})^2 \\ &= \sum_{j=1}^K N_j (\bar{Y}_j - \bar{Y})^2 \end{aligned} \quad (29)$$

Nas equações acima, SS_D calcula o quanto cada amostra difere do seu próprio grupo ($Y_{ij} - \bar{Y}_j$) e SS_E calcula a variação da diferença da média de cada grupo com a média global ($\bar{Y}_j - \bar{Y}$). Para a realização do teste é calculada a estatística F , definida como

$$F = \frac{MS_E}{MS_D}, \quad (30)$$

em que MS_E e MS_D referem-se aos valores das somas dos quadrados divididos por seus respectivos graus de liberdade: SS_E/v_1 e SS_D/v_2 . Como o MS_E calcula a variação da média de cada grupo para a média global, e H_0 tem como hipótese que todas as médias são iguais, espera-se que se H_0 for aceita, o valor de MS_E seja baixo e equivalente à variância dos dados (σ^2), assim como o valor de MS_D , que também será baixo e equivalente à σ^2 . Sendo assim, se o valor da estatística F for próximo de 1, espera-se que H_0 seja aceita (SELTMAN, 2012).

Para determinar se o valor obtido para a estatística F irá rejeitar ou aceitar H_0 , é calculado o valor crítico da pdf F , uma distribuição assimétrica à direita, definida por dois graus de liberdade (v_1 e v_2), sendo $N - K$ e $K - 1$, respectivamente. O valor crítico é obtido para um determinado nível de significância (α), e os valores v_1 e v_2 . A área sobre a curva de F , para valores maiores que o valor crítico, é chamada p-valor. Após a obtenção da estatística F e do p-valor é verificado se H_0 será aceita ou rejeitada, sendo

$$\begin{cases} H_0, & \text{se p-valor} > \alpha \\ H_1, & \text{senão} \end{cases} \quad (31)$$

Ao usar o ANOVA para testar a igualdade da média de pelo menos três grupos, não é possível identificar quais grupos diferem. Portanto, para analisar quais diferem, foi aplicado o teste *post-hoc* (*a posteriori*) Tukey HSD.

2.11.3 Tukey HSD

O teste *post-hoc* Tukey HSD foi utilizado neste trabalho após a aplicação do ANOVA apenas para casos em que a hipótese H_0 foi rejeitada. Por possuir taxa de erro exata de α quando os grupos possuem o mesmo tamanho, o Tukey HSD é um dos testes estatísticos mais utilizados para a comparação de pares de médias (duas a duas) (ABDI; WILLIAMS, 2010). O teste calcula a diferença honestamente significativa entre as médias de dois grupos a partir dos valores de MS_D obtidos da ANOVA e de seus respectivos graus de liberdade ($N - K$), sendo

$$HSD = q_{k,\alpha,gl} \sqrt{\frac{MS_D}{n_k}}, \quad (32)$$

em que q_{k,α,g_l} corresponde a distribuição da amplitude na forma t *Student q*, definida para k grupos, com intervalo de significância α e $g_l = N - K$ graus de liberdade, e n_k corresponde à quantidade de amostras por grupo. Por fim, a partir do valor de *HSD* obtido, duas médias são estatisticamente diferentes se o valor absoluto da diferença entre elas for maior ou igual ao valor da diferença honestamente significativa, ou seja, maior ou igual do que o valor *HSD*.

2.12 Redes neurais convolucionais

As redes neurais convolucionais (*Convolutional neural networks* — CNNs) vêm se popularizando (KHOVOSTIKOV et al., 2018) na área de processamento de imagens médicas, incluindo aplicações voltadas para a detecção da DA (YANG; RANGARAJAN; RANKA, 2018) e estimação da idade do cérebro (DINSDALE et al., 2021; COLE et al., 2017; PENG et al., 2021; ITO et al., 2018; HUANG et al., 2017; UEDA et al., 2019). As CNNs são redes do tipo *feed-forward*, possuem múltiplas camadas e, em geral, são desenvolvidas para reconhecer características de interesse em imagens 2D com alto grau de invariância à translação, escalonamento, distorção, entre outras transformações (LECUN; BENGIO et al., 1995).

2.12.1 Arquiteturas

A primeira rede CNN, proposta em 1998, é denominada *LeNet* (LECUN et al., 1998). Ela foi desenvolvida com o intuito de auxiliar no reconhecimento de caracteres da base MNIST⁴ e gerou resultados com alta acurácia. Na CNN, ao invés de aprender os pesos entre neurônios totalmente conectados, são aprendidos os pesos de filtros que realizam a operação de convolução. Dessa maneira, além de manter a relação espacial das imagens, geralmente 2D, são extraídas características relevantes durante a aprendizagem. Além disso, devido ao uso de mais camadas convolucionais do que as totalmente conectadas, há menor quantidade de parâmetros para serem estimados.

A arquitetura introduzida pela *LeNet* é formada por uma sequência de camadas, sendo que cada uma delas possui uma função específica na propagação do

⁴ <<http://yann.lecun.com/exdb/mnist/>>

sinal de entrada. Tradicionalmente, existem três tipos principais de camadas: (i) as convolucionais; (ii) as de subamostragem (ou *pooling*); e (iii) as totalmente conectadas (*fully-connected*). As primeiras são responsáveis por extrair características das camadas anteriores ou da imagem de entrada, caso a seja a primeira camada da rede, a segunda por reduzir a dimensionalidade dos dados, além de ajudar a tornar a representação invariante a pequenas translações na entrada. Por último, a totalmente conectada é responsável pela propagação do sinal por meio da multiplicação ponto a ponto e o uso de uma função de ativação. A última camada tem como valor de saída a probabilidade de uma imagem de entrada pertencer a uma das classes para qual a rede foi treinada.

Em 2012, alguns anos após a introdução da *LeNet*, devido ao aumento de recursos computacionais e dados disponíveis, surgiu uma das primeiras arquiteturas que revolucionou o uso das CNNs, a AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Ela foi a vencedora da competição ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*) 2012 e deu início a uma sequência de novas propostas de arquiteturas de CNNs com diferentes conceitos, melhorias e aplicações. A seguir, uma breve descrição das particularidades e motivações para a criação das novas arquiteturas é fornecida. Foram descritas apenas as redes mais utilizadas na literatura que contém os conceitos necessários para auxiliar no entendimento deste trabalho:

AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) possui arquitetura similar à *LeNet*, com a inclusão de mais camadas, filtros, empilhamento das convoluções e usou pela primeira vez a função de ativação ReLu. A rede possui filtros de resolução 11×11 , 5×5 e 3×3 , camadas de *pooling* e totalmente conectadas, totalizando 8 camadas com cerca de 61 milhões de parâmetros.

VGG (SIMONYAN; ZISSERMAN, 2014) ficou em segundo lugar no ILSVRC 2014. Surgiu com a proposta de criar uma arquitetura de rede mais profunda que as anteriores. A rede contém 19 camadas e 138 milhões de parâmetros. Os autores conseguiram aumentar a quantidade de camadas com o empilhamento de filtros 3×3 ; eles mostraram que o empilhamento de três desses filtros (243 parâmetros) tinham o mesmo campo receptivo do que um filtro de 7×7 (3969 parâmetros). Dessa maneira, foi possível criar uma rede mais profunda sem aumentar muito a quantidade de parâmetros, permitindo, assim, o aprendizado de

mais não-linearidades. Existem outras versões da rede com variações principalmente na quantidade de camadas, como a VGG-11, 13 e 16.

GoogleNet (SZEGEDY et al., 2015) foi a vencedora do ILSVRC 2014. A rede possui 22 camadas, e pelo fato de não utilizar camadas totalmente conectadas, possui apenas 4 milhões de parâmetros. Também com a proposta de inserção de mais camadas, foi criado o conceito de *Inception Module*. Além disso, ao invés de escolher quais tamanhos de filtros são os mais adequados, sua arquitetura executa todos em paralelo, como 1×1 , 3×3 , 5×5 e *max pooling* de 3×3 , e empilha as respostas. Dessa maneira, a rede aprende quais são as melhores configurações de filtros, dando maior dinamicidade. No entanto, ao usar todos os filtros, a quantidade de parâmetros da rede aumenta, inviabilizando que seja mais profunda. Para contornar esse problema, são inseridas camadas chamadas de “*bottleneck*”, que adicionam filtros de tamanho 1×1 antes da convolução desejada, cuja quantidade de filtros seja inferior à que se deseja na camada subsequente de convolução. Para camadas de *max pooling*, a camada *bottleneck* é inserida após a operação. Com isso, os filtros são executados em paralelo com custo bastante reduzido.

ResNet (HE et al., 2016) vencedora do ILSVRC 2015. A rede possui 152 camadas e foi a primeira a obter erro menor do que o humano. Sua principal inovação foi conseguir criar redes ultra profundas sem sacrificar seu desempenho. Os autores notaram que redes mais profundas enfrentavam o “problema da degradação”, ou seja, a acurácia da rede acaba saturando ou degradando com o aumento na quantidade de camadas, indicando que redes mais rasas estão aprendendo melhor do que mais profundas. Esse problema pode ser causado pelo desaparecimento do gradiente, em que o gradiente pode ser tornar infinitamente pequeno na retro propagação para camadas as anteriores e/ou maldição da dimensionalidade. Sabendo que redes mais rasas têm melhor desempenho do que as mais profundas com algumas camadas a mais adicionadas, os autores propuseram pular essas camadas extras para obter pelo menos a mesma acurácia que a rede mais rasa. Para isso, foi criada a “conexão identidade”, que insere uma conexão entre o início (x) e final ($H(x)$) de um conjunto de camadas ($F(x)$), sendo estas camadas conhecidas como o mapeamento residual. Dessa maneira, $H(x)$ é definido pelo somatório de $F(x)$ e x , permitindo que algumas camadas sejam puladas durante o treinamento. A função identidade pode ser aprendida simplesmente definindo o residual como zero.

Além disso, devido aos saltos nas conexões, a rede pode escolher quais camadas são mais importantes, tornando-a mais dinâmica e o gradiente pode ser propagado para as camadas iniciais, permitindo a criação de redes ultra profundas.

Até então, o enfoque das arquiteturas foi em aumentar a acurácia dos modelos com redes mais profundas. A seguir são brevemente descritas arquiteturas que focam também em desempenho e eficiência.

MobileNet — V1 (HOWARD et al., 2017) as *MobileNets* foram criadas para funcionar em tempo hábil e em uma plataforma limitada computacionalmente em aplicações como robótica, carro autônomo, realidade aumentada e *mobile*. Sua primeira versão, v1, propôs o uso de blocos chamados convolução *depthwise separable* ao invés da convolução tradicional. Esses blocos possuem resoluções e tamanhos de entrada e saída correspondentes à convolução tradicional, porém com redução significativa da quantidade de operações e custo computacional, permitindo que sejam utilizados em ambiente com maiores limitações computacionais. A convolução *depthwise separable* possui duas etapas: a convolução *depthwise* e a convolução *point-wise*. A primeira consiste na realização de uma convolução separada para cada canal (*channel*) de cor da imagem e posterior concatenação das respostas, e a segunda, que possui como entrada a saída da primeira, consiste na realização de convoluções de $1 \times 1 \times N_c$ para cada quantidade de filtros.

MobileNet — V2 (SANDLER et al., 2018) a segunda versão é composta por blocos *mobile inverted bottleneck convolution* (MBConv) cujo objetivo é permitir que a rede aprenda funções mais complexas enquanto mantém a quantidade de memória relativamente pequena, além de possuir resoluções e tamanhos de entrada e saída correspondentes à convolução tradicional. Cada bloco é delimitado por uma “conexão identidade”, similar à *ResNet* (HE et al., 2016), sendo as camadas intermediárias denominadas de mapeamento residual. Esse mapeamento contém uma camada de expansão, a convolução *depthwise* e a convolução *point-wise* (denominada projeção na v2). A operação de expansão aumenta o tamanho da representação no bloco por um determinado fator, permitindo que a rede aprenda funções mais ricas, a convolução *depthwise* realiza a convolução por canal e a projeção retorna ao tamanho desejado para haver economia de memória.

Squeeze and excitation (HU; SHEN; SUN, 2018) essa rede propôs o uso de blocos *squeeze and excitation* para recalibrar de maneira adaptativa as res-

postas por canal, modelando explicitamente suas interdependências. Sua criação foi motivada devido à operação de convolucional tradicional unir as informações espaciais e por canal, ao invés de analisar de maneira separada. A parte espacial foi explorada em (BELL et al., 2016; NEWELL; YANG; DENG, 2016), deixando a análise do relacionamento dos canais em aberto. Um bloco é composto por duas etapas, a *squeeze* e a *excitation*. A camada *squeeze* consiste na operação de *max pooling* por canal, gerando um vetor de tamanho $1 \times 1 \times N_c$. A etapa de *excitation* possui quatro camadas sequenciais, sendo uma totalmente conectada de tamanho $1 \times 1 \times N_c/r$, uma função de ativação ReLu, totalmente conectada de tamanho $1 \times 1 \times N_c$ e uma função de ativação sigmoide, em que r é um hiperparâmetro. Dessa maneira, o bloco consegue suprir características menos informativas e enfatizar as características mais informativas. Embora as camadas totalmente conectadas aumentem a quantidade de parâmetros, a inserção dos blocos produz ganhos significativos de acurácia.

***EfficientNets* (TAN; LE, 2019)** As *EfficientNets* correspondem a uma família de modelos de CNNs (2D) criadas pelo *Google research brain team*⁵ que atingiram acurácia e eficiência melhores na base do *ImageNet* do que as CNNs propostas anteriormente. Sua principal motivação foi construir redes com alta acurácia e eficiência, pois as arquiteturas de rede existentes são comumente aumentadas para obter melhor acurácia quando mais recursos são disponibilizados. A prática convencional para o dimensionamento do modelo é aumentar arbitrariamente a profundidade ou largura da CNN, ou usar uma imagem de entrada com resolução maior. Embora esses métodos melhorem a acurácia, geralmente exigem ajuste manual e, ainda assim, costumam produzir um desempenho abaixo do ideal. A proposta da *EfficientNet* é escalar automaticamente a arquitetura das redes para uma aplicação ou dispositivos específicos. Essa família de redes é baseada em uma arquitetura base, a B0, a qual foi criada a partir de uma pesquisa de arquiteturas que otimizam, simultaneamente, a acurácia e as operações de ponto flutuante por segundo (*Floating point operations per second* — FLOPS). Sua arquitetura é composta principalmente por blocos MBConv (SANDLER et al., 2018), com a adição da otimização *squeeze and excitation* (HU; SHEN; SUN, 2018). A partir da arquitetura base são realizadas aplicações de dimensionamento

⁵ <https://research.google/teams/brain/>

que alteram a profundidade ($\#layers$), largura ($\#channels$) e resolução da rede simultaneamente.

2.12.2 Taxa de aprendizagem cíclica

Para o treinamento das redes neurais, optamos pela taxa de aprendizagem cíclica (*Cyclical Learning Rate* — CLR) (Smith, 2017). Existem diversas políticas utilizadas para lidar com a taxa de aprendizado (TA) durante o treinamento. As mais utilizadas incluem o decaimento da TA após algumas épocas (*StepLR*⁶) ou quando a função de perda parar de melhorar (*ReduceLROnPlateau*⁷). A CLR varia a TA entre um limite inferior e superior (*base_lr* e *max_lr*) usando uma política de decaimento. Na prática, essas mudanças periódicas nos valores da TA ajudam a evitar pontos de sela ou mínimos locais, conseqüentemente acelerando o processo de treinamento.

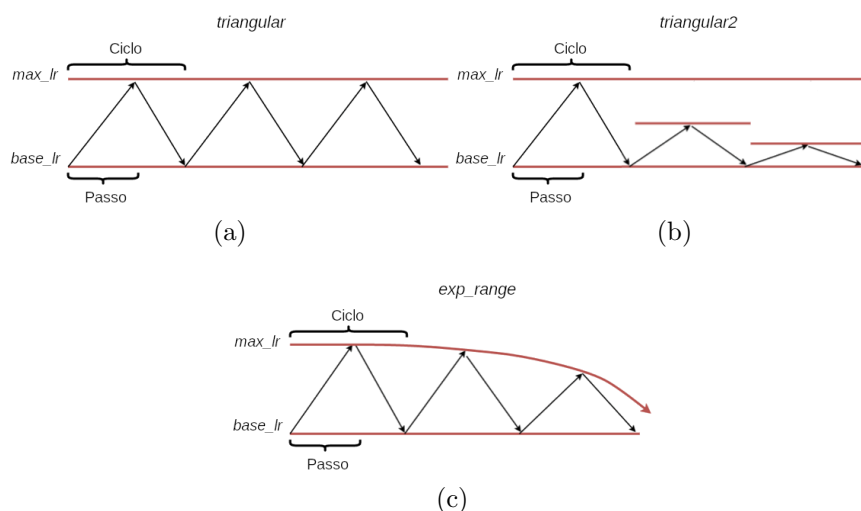


Figura 5 – Política de decaimento da CLR.

A CLR altera o valor da TA após cada iteração (execução de um *batch*) e de acordo com uma política de decaimento — ilustradas na Figura 5. As políticas podem ser *triangular* (a), *triangular2* (b) e *exp_range* (c). Todas seguem um

⁶ https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.StepLR.html

⁷ https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html#torch.optim.lr_sche

padrão triangular, que se inicia com um valor baixo (*base_lr*), aumenta até um valor máximo (*max_lr*) e depois retorna para o valor baixo. Nas políticas *triangular2* e *exp_range*, o valor máximo cai pela metade e por um fator exponencial, respectivamente. O padrão é seguido durante todo o treinamento.

2.13 Métricas de avaliação

Nesta seção são apresentadas as métricas utilizadas para avaliar os resultados de classificação e regressão das imagens e a análise estatística dos atributos extraídos.

2.13.1 Classificação

Para avaliar a classificação das imagens de RM, foram utilizadas a área sob a curva ROC (AUC), acurácia (Acc), medida F1 (F1), sensibilidade (Sens) e especificidade (Espec). A definição formal de cada métrica é apresentada a seguir, em que VP representa os verdadeiros positivos, FP os falsos positivos, VN representa os verdadeiros negativos e FN os falsos negativos. A quantidade de ocorrências de cada valor compõe a matriz de confusão (ou contingência), indicada na Tabela 1.

Tabela 1 – Matriz de Confusão

		Predita	
		Positiva	Negativa
Verdadeira	Positiva	TP	FN
	Negativa	FP	TN

A partir da Tabela 1, são calculadas duas métricas que se complementam, a sensibilidade e especificidade. A sensibilidade calcula a porcentagem de acertos da classe positiva e a especificidade calcula a porcentagem de acertos da classe negativa, sendo

$$\text{sensibilidade} = \frac{VP}{VP+FN}, \quad (33)$$

$$\text{especificidade} = \frac{VN}{VN+FP}. \quad (34)$$

A partir dos valores de sensibilidade e especificidade, é possível calcular a acurácia balanceada, que para o caso binário, corresponde a média entre os dois valores (KELLEHER; NAMEE; D'ARCY, 2015). Da Tabela 1 obtém-se também a acurácia, que calcula a porcentagem de acertos em relação ao total, e a medida F1, sendo a média harmônica das medidas de precisão e sensibilidade, calculadas como

$$\text{acurácia} = \frac{VP+VN}{VP+FP+VN+FN}, \quad (35)$$

$$F1 = \frac{2VP}{2VP+FP+FN}. \quad (36)$$

Além disso, a curva ROC, que consiste em um gráfico que mostra o desempenho do modelo de classificação, pode ser determinada variando-se os limiares de classificação utilizados para a construção da Tabela 1. No eixo x são plotados os valores de 1-especificidade e no eixo y os valores de sensibilidade. A AUC, sendo a área sob a curva ROC, fornece uma medida de desempenho global de um classificador.

2.13.2 Regressão

Para avaliar o modelo de regressão, foram utilizadas o erro médio (*Mean error* — ME), o erro médio absoluto (*Mean absolute error* — MAE) e a raiz quadrada do erro médio quadrático (*Root mean squared error* — RMSE) e a correlação de Pearson. A definição formal de cada métrica é apresentada a seguir.

As métricas utilizadas, ME, MAE e RMSE são definidas com base no erro de predição, definidas como:

$$ME = \frac{1}{n} \sum_{i=1}^n \hat{y}_i - y_i, \quad (37)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (38)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2, \quad (39)$$

$$RMSE = \sqrt{MSE}, \quad (40)$$

em que y_i corresponde ao valor esperado (rótulo real) \hat{y}_i ao valor predito.

O coeficiente de correlação de Pearson, r , visa mensurar o grau de relacionamento e o sinal da correlação linear entre duas variáveis. É formalmente definida como:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}},$$

em que r varia entre $[-1, 1]$; o valor positivo indica que as variáveis se relacionam de maneira diretamente proporcional e o negativo que se relacionam de maneira inversamente proporcional. Independente do sinal, quanto mais próximo de 1, maior a correlação, ao passo que, quanto mais próximo de 0, menor a correlação. Os valores de x e y , correspondem às duas variáveis analisadas.

2.14 Considerações finais

Este capítulo apresentou informações gerais sobre o desenvolvimento da doença de Alzheimer e o processo de formação das imagens de RM. Além disso, forneceu informações mais detalhadas sobre os algoritmos de pré-processamento, o banco de filtros log-Gabor 3D, a técnica de congruência de fase 3D, a criação do atlas probabilístico de pontos salientes e a segmentação das imagens. Forneceu também explicações sobre testes estatísticos, o classificador SVM, arquiteturas de redes neurais convolucionais e as métricas de avaliação. Os conceitos apresentados servem como embasamento teórico para o entendimento desta pesquisa. No capítulo seguinte, é apresentada uma descrição de trabalhos propostos na literatura que possuem intersecções conceituais e práticas com esta pesquisa.

Capítulo 3

Revisão Bibliográfica

Este capítulo apresenta uma revisão dos principais trabalhos propostos na literatura para a classificação da DA utilizando atributos de atrofia estrutural, os que avaliaram os efeitos da assimetria estrutural na DA, assim como trabalhos que objetivam prever a idade biológica do cérebro.

3.1 Atrofias estruturais

A atrofia estrutural, um dos principais enfoques desta pesquisa, é considerada um importante biomarcador da DA, pelo fato de ser parte integrante de seu processo fisiopatológico. A DA é causada pela lesão neuronal e neuro degeneração progressiva do cérebro, sendo os hipocampus, estruturas essenciais na formação de memórias, um dos primeiros a atrofiar à medida que mais neurônios morrem. Portanto, a análise de alterações hipocampais é o motivo principal de vários estudos, visando à determinação de um biomarcador para a doença.

Existe uma variedade de técnicas para o auxílio ao diagnóstico da DA presentes na literatura. Alguns trabalhos utilizam técnicas sem intersecção com as utilizadas e propostas nesta pesquisa e podem se distanciar do nosso enfoque, portanto, são apenas brevemente descritos a seguir para mostrar a diversidade das propostas.

Os métodos presentes podem ser divididos por categoria de análise, como focados em voxels, ROIs e *patches*, e por metodologia de extração de atributos, como engenharia ou aprendizado profundo.

Métodos que se baseiam voxels, em geral, usam mapas de densidade de GM e WM gerados por morfometria baseada em *voxel* (*Voxel based morphometry* — VBM) como atributos para classificação. Klöppel et al. (2008), por exemplo, utilizaram mapas de densidade de GM extraídos de todo o encéfalo como atributos de entrada para um classificador SVM para auxiliar no diagnóstico da DA. Existem também, abordagens que incluem algoritmos de seleção/redução de atributos, visto que as imagens volumétricas de RM possuem milhões de *voxels*. Nesse caso, Salvatore et al. (2015) utilizaram a análise dos componentes principais (*Principal component analysis* — PCA) para reduzir a dimensão de mapas de densidade de GM e WM e um classificador SVM para realizar a classificação da DA e prever a conversão do CCL para a DA.

Métodos com enfoque em ROIs geralmente usam estruturas do cérebro definidas com base no histórico da literatura médica, como o hipocampo e o córtex entorrinal. Ahmed et al. (2015b) utilizaram atributos extraídos dos hipocampus com funções harmônicas circulares (TOEWS et al., 2010) e atributos do CSF para realizar a classificação da DA. Cada conjunto de atributos foi inserido em diferentes classificadores SVM, cujas saídas foram usadas como entrada em outro classificador para refinar os resultados. Sørensen et al. (2016) extraíram atributos de textura das regiões dos hipocampus com o objetivo de mostrar que as alterações dos padrões de textura das imagens podem ser mais sensíveis aos estágios iniciais da DA do que alterações nos volumes hipocampais. Os atributos foram inseridos em um classificador SVM para distinguir indivíduos com CCL quando comparados aos que são CN e possuem DA. Coupé et al. (2012b) utilizaram atributos dos hipocampus e córtex entorrinais produzidos pelo algoritmo *Scoring by nonlocal image patch estimator* (SNIPE) (COUPÉ et al., 2012a) e treinaram o classificador de análise de discriminantes lineares (*Linear discriminant analysis* — LDA) para o diagnóstico da DA.

Dependendo da quantidade de ROIs, também são utilizados algoritmos de seleção/redução de atributos. Assim, Eskildsen et al. (2015) utilizaram o algoritmo *Mutual Information Feature Selection* (MIFS) para selecionar os atributos dos hi-

pocampos e córtex entorrinais, produzidos com o SNIPE, e atributos de espessura cortical, produzidos com a *Fast Accurate Cortex Extraction* (ESKILDSEN; ØSTERGAARD, 2006) e treinaram um classificador LDA para prever a conversão do CCL para a DA.

É importante mencionar que as divisões de metodologia mencionadas não são restritas, existem trabalhos que realizam análises com diversas combinações. Métodos que utilizam *patches* são frequentemente encontrados com metodologias de *voxels*, ROIs e seleção de atributos, além de, em alguns casos, utilizarem pontos salientes (*landmarks*). Esses trabalhos apresentam metodologias que possuem maior intersecção com a proposta desta pesquisa, portanto são descritos com maior detalhamento a seguir.

3.1.1 Trabalhos correlatos

Zhang et al. (2016) propuseram uma metodologia para detecção de *landmarks* e extração de atributos morfológicos para o diagnóstico da DA. Os autores utilizaram 1619 imagens de RM das bases de dados ADNI I-II, sendo 430 CN, 831 CCL e 358 de pacientes com DA. Para selecionar *landmarks* que fossem discriminativos, foram utilizados dois critérios: atributos diferentes entre populações e maior variância local, denominados como AD e ativos, respectivamente.

Para o primeiro foi criado um mapa de p-valor para os *voxels* da imagem de referência Colin27, obtido por comparações entre atributos morfológicos dos *voxels* de dois grupos de diagnóstico, CN e AD das imagens de treinamento (ADNI-I). Um total de 50 atributos foram extraídos utilizando a técnica *Local Energy Pattern* (ZHANG; LIANG; ZHAO, 2012). A partir do mapa de p-valor, os *voxels* estatisticamente significativos foram selecionados e denominados *landmarks* AD.

Para o segundo foi criada uma ponderação entre mapas de saliência e inconsistência, definidos, respectivamente, pelos cálculos de entropia entre *patches* de diversos tamanhos extraídos dos *voxels* de todas as imagens de treino e a imagem de referência e da variância local. Os dois mapas foram ponderados e os valores máximos foram selecionados. As duas categorias de *landmarks*, AD e ativos, foram selecionadas e seus atributos utilizados para criar um mapeamento não linear entre o *patch* ao redor do *voxel* e seu deslocamento 3D para o ponto de destino (apren-

dido com o algoritmo *regression-forest* com restrição de forma). Para as imagens de teste (ADNI-II), os *landmarks* foram detectados usando o modelo estimado no treinamento. Os *landmarks* ativos foram detectados usando o modelo *regression-forest* com restrição de forma e os *landmarks* AD foram detectados utilizando os *landmarks* ativos como referência.

Para a classificação foram utilizados atributos morfológicos extraídos da vizinhança de cada *landmark* AD de treinamento e teste. Nesse caso, os atributos de cada *landmark* foram concatenados horizontalmente e utilizados como entrada para um classificador SVM linear com validação cruzada em 2-folds. Os autores detectaram 1741 *landmarks* para o primeiro e 1761 para o segundo *fold*. Como resultados, eles obtiveram uma distância média de 2,41 mm de erro para a detecção de *landmarks* e os seguintes valores para classificação CN×DA: 83,7% de acurácia, 83,6 de acurácia balanceada, 80,5% de sensibilidade e 85,1% de especificidade. Já para a classificação de CN×CCL os resultados obtidos foram: 73,6% de acurácia, 72,5 de acurácia balanceada, 75,3% de sensibilidade e 69,7% de especificidade.

Embora esse trabalho tenha sido desenvolvido para a detecção de *landmarks*, os autores obtiveram bons resultados de classificação, com a ressalva de que os atributos de cada *landmark* foram concatenados horizontalmente, o que pode ser problemático, devido à alta dimensionalidade do espaço de atributos gerado.

Dando continuidade ao trabalho, Zhang et al. (2017) propuseram a extração de atributos de *landmarks* para o diagnóstico da DA utilizando informações espaciais e longitudinais, ou seja, mais de uma imagem por paciente obtida em diferentes períodos temporais. Nesse caso, a quantidade de imagens longitudinais variou entre três e seis por paciente, com intervalo mínimo de 6 meses entre os exames. Os autores utilizaram 707 imagens de RM da base de dados ADNI I, sendo 154 pacientes com DA, 207 CN e 346 com CCL.

A detecção de *landmarks* foi realizada com a técnica desenvolvida em seu trabalho anterior (ZHANG et al., 2016) e os atributos espaciais foram extraídos a partir de cada *landmark* com a técnica de histograma de gradientes orientados (*Histogram of oriented gradients* — HOG) (DALAL; TRIGGS, 2005) combinada com a estratégia de *bag-of-words* para a inserção de informações das imagens longitudinais. Os atributos longitudinais consistem de HOGs com 72 bins de tamanho,

extraídos de mapas jacobianos construídos considerando o campo de deformação médio entre as imagens de um mesmo paciente e a imagem de referência Colin.

Na etapa de classificação foram utilizadas as características espaciais e longitudinais de cada *landmark*, totalizando 122 atributos e cerca de 1500 *landmarks*. Para cada imagem, os atributos de todos os *landmarks* foram concatenados e utilizados como entrada para um classificador SVM com validação cruzada em 10 *folds*. Como resultados, os autores obtiveram 88,3% de acurácia, 0,94 de AUC, 79,61% de sensibilidade e 94,69% de especificidade para a classificação de CN×DA, e 79,02% de acurácia, 0,852 de AUC, 90,46% de sensibilidade e 59,9% de especificidade para a classificação de CN×CCL.

Embora os autores relataram que a inserção de informações longitudinais trouxe um ganho médio de 2% a 6% em relação à acurácia, essa metodologia ainda não corrigiu o problema que a concatenação horizontal dos atributos pode trazer, visto que a quantidade de *landmarks* detectados também grande. Além disso, os valores de especificidade e sensibilidade obtidos para o experimento CN×CCL são bem distintos, principalmente ao considerar que existe um leve desbalanceamento entre essas classes.

No ano seguinte, Liu et al. (2018) propuseram a utilização de *landmarks* no contexto de *deep learning* com a utilização de CNNs. Os autores utilizaram 1526 imagens de RM das bases de dados ADNI I-II e MIRIAD. A base ADNI I contém imagens distribuídas em 199 DA, 229 CN, 167 CCL progressivos (pCCL), e 226 estáveis (sCCL), enquanto a base ADNI II contém 159 DA, 200 CN, 38 pCCL, e 239 sCCL. Por fim, a base MIRIAD contém 46 DA e 69 CN.

Os *landmarks* foram detectados com a técnica desenvolvida em seu trabalho anterior (ZHANG et al., 2016), dos quais foram extraídos *patches* de tamanho $24 \times 24 \times 24$ *voxels* ao redor de cada *landmark*. Para a organização dos dados foi utilizado o conceito de *bag-of-words*, considerando cada imagem como uma *bag* que contém um conjunto de *patches* e um rótulo. Como nem todos os *patches* extraídos de uma imagem podem ser significativos para representar sua classe, os autores utilizaram aprendizado por múltiplas instâncias (MARON; LOZANO-PÉREZ, 1998; DIETTERICH; LATHROP; LOZANO-PÉREZ, 1997). Nesse caso, para aprender as representações dos *patches* de uma *bag*, os *patches* foram inseridos

paralelamente em várias (quantidade de *patches* entre 30 e 50) arquiteturas CNNs, compostas por seis camadas convolucionais e duas camadas totalmente conectadas. Embora a arquitetura de cada CNN seja a mesma, os parâmetros aprendidos por cada uma são distintos, para possibilitar o aprendizado de características locais e específicas em cada *patch*. Por fim, com o objetivo de extrair características globais, a saída das CNNs foram concatenadas horizontalmente e três outras camadas totalmente conectadas foram inseridas no modelo.

As imagens da base de dados ADNI-I foram utilizadas para o treinamento e as imagens das bases de dados ADNI-II e MIRIAD para teste. Como resultado da classificação das imagens da base de dados ADNI-II em CN×DA, os autores obtiveram 91,09% de acurácia, 0,959 de AUC, 88,05% de sensibilidade e 95,3% de especificidade. Já para a base de dados MIRIAD, os resultados obtidos foram de 92,75% de acurácia, 0,972 de AUC, 93,48% de sensibilidade e 91,3% de especificidade. Para a classificação das imagens em pCCL×sCCL, os autores obtiveram 76,9% de acurácia, 0,776 de AUC, 42,11% de sensibilidade e 82,43% de especificidade para a base de dados ADNI-II.

Pelo fato desse trabalho utilizar *deep learning*, a escolha da utilização de *patches* aumenta significativamente a quantidade de amostras para o treinamento. Entretanto, como os *patches* são combinados em uma mesma arquitetura, a quantidade de imagens utilizadas para o treinamento (127 CN, 199 DA, 167 pCCL e 226 sCCL) ainda pode ser considerada pequena. Além disso, os valores de especificidade e sensibilidade obtidos para o experimento pCCL×sCCL são bastante distintos, principalmente ao considerar a discrepância na quantidade de imagens no teste (38 pCCL vs. 239 sCCL).

A seguir, Lian et al. (2020), propuseram a criação de uma CNN hierárquica para a classificação da DA. Foram utilizadas as mesmas imagens de RM das bases de dados ADNI I-II de seu trabalho anterior (LIU et al., 2018). A organização hierárquica da rede foi realizada usando *patches* no primeiro nível, seguida por regiões no segundo nível e imagem no último. A localização dos *patches* foi definida com base nas localizações correspondentes da imagem de referência Colin27 e os *patches* foram extraídos ao redor dos *landmarks*, detectados com a técnica proposta em (ZHANG et al., 2016).

No primeiro nível da rede, os *patches*, de tamanho $25 \times 25 \times 25$ *voxels*, foram inseridos em 120 (quantidade de *patches*) arquiteturas CNNs. Cada CNN é composta por seis camadas convolucionais que compartilham os mesmos pesos, seguida por uma camada de classificação. Dessa maneira, a capacidade discriminativa de cada *patch* recebe seu valor de acurácia. Em seguida, no segundo nível, a saída dos *patches* foram agrupadas (concatenadas) por regiões, considerando as suas localizações espaciais. Então, para cada região, foi adicionada uma camada convolucional, seguida de uma camada de classificação. Por fim, no nível da imagem, a saída das regiões, em conjunto com seus valores de acurácia, foram concatenadas e inseridas em uma camada convolucional para a classificação da imagem. Durante o treinamento da rede, os *patches* e as regiões foram ranqueados de acordo com seu poder preditivo (acurácia) e apenas os que apresentaram valores maiores que um determinado limiar foram mantidos. Além disso, os autores utilizaram *Data Augmentation* (PEREZ; WANG, 2017) com as operações de rotações, distorções e deslocamentos.

As imagens da base de dados ADNI-I foram utilizadas para treinamento e as imagens das bases de dados ADNI-II para o teste. Como resultados da classificação das imagens em CN \times DA, os autores obtiveram 90,03% de acurácia, 0,951 de AUC, 82,4% de sensibilidade e 96,5% de especificidade. Para a classificação das imagens em pCCL \times sCCL, os autores utilizaram a rede pré-treinada para CN \times DA e aplicaram a técnica de *Transfer Learning* (TAJBAKHSI et al., 2016). Nesse experimento, os autores obtiveram 80,09% de acurácia, 0,781 de AUC, 52,6% de sensibilidade e 85,4% de especificidade.

Pelo fato da quantidade de imagens ser pequena, os autores utilizaram estratégias de *data augmentation* e conseguiram um pequeno aumento em seus resultados em relação ao seu trabalho anterior e ao estado da arte. No entanto, os valores de especificidade e sensibilidade obtidos para o experimento pCCL \times sCCL ainda são muito distintos, principalmente por existir uma grande discrepância na quantidade de imagens no conjunto de teste (38 pCCL vs. 239 sCCL). Além disso, os autores mencionam que selecionam os *patches* e as regiões com base nos que obtiveram maior acurácia, o que pode ser problemático dado o desbalanceamento da base de dados.

A Tabela 2 apresenta um sumário dos principais trabalhos publicados na lite-

ratura em classificação das imagens RM estruturais para a DA.

Todos os trabalhos relatados apresentaram desempenhos satisfatórios em seus experimentos, com metodologias bastante consistentes e com bons resultados de classificação. No entanto, cabe ressaltar algumas limitações encontradas: a falta de experimentos que explorem a classificação CCL \times DA, ausência de indicação e discussão dos desvios-padrão obtidos entre os *folds*, grande desbalanceamento entre classes e a utilização de atributos que podem não representar alterações cerebrais causadas pela DA e pelo CCL, que conseqüentemente, podem gerar atributos com alta dimensionalidade.

Com base na discussão anterior, para o desenvolvimento deste projeto buscou-se superar as principais limitações das diferentes técnicas apresentadas na literatura. De maneira resumida, este estudo apresenta uma nova metodologia baseada em ROIs e *landmarks* para a classificação de imagens de RM nas classes DA e CCL. São realizadas análises em dois níveis, o de *landmarks* e o da imagem, evitando assim vetores de atributos com alta dimensão. No primeiro nível, *landmarks* discriminativos das regiões hipocampais são automaticamente detectados e selecionados usando um atlas de pontos salientes como referência. Essa abordagem reduz o número de *landmarks* avaliados, visto que são removidos pontos instáveis ou ruidosos, diminuindo, portanto, o número de comparações. Além disso, a análise em dois níveis permite um aumento do número de amostras no primeiro nível e geração de um vetor de atributos muito pequeno e interpretável para o segundo nível. A metodologia desenvolvida para a realização desta análise é apresentada no próximo capítulo.

Tabela 2 – Resumo dos principais trabalhos publicados na literatura relacionados ao tema desta pesquisa.

Trabalho	# Pacientes			CN×CCL				CCL×DA				CN×DA			
	CN	MCI	AD	Acc(%)	Sens(%)	Spec(%)	AUC	Acc(%)	Sens(%)	Spec(%)	AUC	Acc(%)	Sens(%)	Spec(%)	AUC
Liu et al. (2014)	77	169	65	76,92	74,29	78,13	-					87,76	88,57	87,22	-
Ye et al. (2016)	52	99	51	68,02	76,87	51,15	0,71	-	-	-	-	87,26	88,43	86,15	0,93
Zhang et al. (2016)	430	731	358	73,6	75,3	69,7	-	-	-	-	-	83,7	80,5	85,1	-
Sørensen et al. (2016)	140	233	101	-	-	-	0,806	-	-	-	-	-	-	-	0,915
Korolev et al. (2017)	61	43 (pMCI)	50	63,00	-	-	0,67	62,00	-	-	0,61	79,0	-	-	0,88
Zhang et al. (2017)	207	346	154	79,02	90,46	59,9	0,85					88,3	79,61	94,69	0,94
Cao et al. (2017)	229	397	192	71,9	79,0	66,7	0,73	74,7	69,1	79,1	0,76	88,6	85,7	90,4	0,898
Liu et al. (2018)	429	-	358	-	-	-	-	-	-	-	-	91,09	88,05	95,3	0,959
Liu et al. (2019)	249	-	227	-	-	-	-	-	-	-	-	93,7	94,6	93,2	0,986
Lian et al. (2020)	429	-	358	-	-	-	-	-	-	-	-	90,3	82,4	96,5	0,951
Liu et al. (2020)	119	233	97	76,2	79,5	69,8	0,775					88,9	86,6	90,8	0,925

3.2 Assimetrias estruturais

A assimetria inter hemisférica estrutural do cérebro, explorada por esta pesquisa, é outro aspecto importante a ser considerado na análise do desenvolvimento da DA. Os dois hemisférios do cérebro diferem em anatomia e função, acarretando em diferenças assimétricas inter hemisféricas que podem estar relacionadas a diversos fatores, como percepção auditiva, preferências motoras, escrita, acuidade sensorial, gênero, idade, fatores genéticos e influências hormonais (TOGA; THOMPSON, 2003). Além de possíveis assimetrias devido a fatores funcionais, alterações na assimetria cerebral também ocorrem devido ao envelhecimento normal (com menor intensidade) ou a processos neurodegenerativos patológicos (com maior intensidade), portanto, a análise de tais assimetrias pode auxiliar na identificação do início de uma condição neurológica precoce (TEVEROVSKIY et al., 2008; SHI et al., 2009; SARICA et al., 2018; ARDEKANI et al., 2019).

Nos últimos anos, estudos avaliaram as assimetrias hemisféricas nas imagens de RM, concentrando-se principalmente em populações saudáveis estratificadas por gênero (ASHBURNER et al., 1998; GOOD et al., 2001; LUDERS et al., 2004; LUCARELLI et al., 2013; KOVALEV; KRUGGEL; CRAMON, 2002) e destreza (GOOD et al., 2001; HERVÉ et al., 2006). Outros avaliaram a assimetria cerebral sob a influência de doenças cognitivas, como autismo e distúrbio da linguagem do desenvolvimento (HERBERT et al., 2005), epilepsia (MARTINS et al., 2019; MECHANIC-HAMILTON et al., 2009), esquizofrenia (HECKERS; KONRADI, 2010), DA (SARICA et al., 2018; ARDEKANI et al., 2019; BARNES et al., 2005; THOMPSON et al., 2003; SPERLING, 2007) e doença de Parkinson com análise das imagens do DaTSCAN (ORTIZ et al., 2019). Na subseção a seguir são descritos alguns estudos que exploraram a assimetria cerebral com principal enfoque na DA, visando estabelecer os principais achados e contribuições sobre o desenvolvimento assimétrico do cérebro na presença dos processos neurodegenerativos da doença.

3.2.1 Trabalhos correlatos

Shi et al. (2009) conduziram uma meta-análise sobre o volume hipocampal em pacientes com CCL e DA. A meta-análise é um método estatístico que, ao invés de

resumir os resultados dos trabalhos, define um *effect size* (medida quantitativa da magnitude de um fenômeno) para representá-los, permitindo que sejam realizadas comparações entre eles. Os autores utilizaram 28 trabalhos publicados a partir de 1998 que realizaram estudos sobre o volume hipocampal em populações com CCL e DA. Dentre eles, 14 analisaram as classes CN e DA, 5 CN e CCL, e 9 analisaram CN, CCL e DA. Todos os estudos possuíam poucas amostras (entre 7 e 103 para as classes CCL e DA), portanto foram integradas com a meta-análise e resultaram em mais de 300 amostras por classe. Além disso, após a obtenção dos dados, os autores também analisaram a lateralidade da assimetria hipocampal. A análise dos dados foi conduzida com o *effect sizes* ponderados, que indicam a magnitude da associação em todos os estudos e a análise da lateralidade foi conduzida com o cálculo de *effect sizes* distintos para os hipocampos esquerdo e direito. Como resultados, os autores encontraram acentuada atrofia bilateral hipocampal nas classes CCL e DA, com redução dos tecidos cerebrais predominantemente do lado esquerdo. Comparando a redução volumétrica inter hemisférica entre as classes CN e CCL, os autores reportaram uma redução média de 12,9% no volume do hipocampo esquerdo e 11,1% do hipocampo direito. Além disso, comparando as classes CCL e DA, os autores encontraram um padrão severo de atrofia bilateral do hipocampo na DA maior que no CCL, com redução média de 24,2% no volume do hipocampo esquerdo e 23,2% do direito.

Lucarelli et al. (2013) analisaram a magnitude da assimetria volumétrica hipocampal em imagens de RM de 1975 pacientes que não apresentavam histórico de derrame ou grandes anomalias estruturais cerebrais da base de dados *Dallas Heart Study* (VICTOR et al., 2004). A segmentação hipocampal foi realizada com os algoritmos FIRST (PATENAUDE et al., 2011) e FreeSurfer¹, com o objetivo de compará-los. A assimetria hipocampal foi calculada pelo valor absoluto da diferença entre os volumes dos hipocampos esquerdo e direito, dividido pelo somatório dos dois volumes. Nas análises estatísticas, os autores verificaram o comportamento de seu índice de assimetria por gênero e o correlacionaram com a idade. Para a comparação entre os gêneros foi utilizado o teste *Mann-Whitney U* (MANN; WHITNEY, 1947) e a correlação etária com o *Speakman rank-order* (GOULDEN, 1939), ambos com nível de significância de 0,05. Como resultados,

¹ <http://surfer.nmr.mgh.harvard.edu/>

os autores reportaram assimetria de 9,8% com o FIRST e 4,5% com o FreeSurfer, além disso, relataram também maior assimetria em homens que mulheres, e um aumento da assimetria conforme o aumento da idade.

Evidências, apresentadas na literatura médica (TOGA; THOMPSON, 2003), e confirmadas pela meta-análise de Shi et al. (2009), indicam que fatores patológicos podem levar a uma assimetria inter hemisférica do cérebro mais acentuada e que a diferença entre os volumes hipocampais existe e varia com a progressão de doenças. No entanto, o grau de assimetria nos subcampos do hipocampo foi pouco investigado. Dessa maneira, Sarica et al. (2018) conduziram um estudo para avaliar a assimetria de subestruturas do hipocampo visando encontrar possíveis biomarcadores para a DA. Os autores utilizaram imagens de 400 pacientes da base ADNI, sendo 100 CN, 100 pCCL, 100 sCCL e 100 com DA. No total foram analisadas 12 estruturas cujas segmentações foram realizadas pelo algoritmo FreeSurfer. O índice de assimetria de cada estrutura foi calculado pelo valor absoluto da diferença entre os volumes esquerdo e direito multiplicados por 100% e dividido pelo somatório dos dois volumes, sendo que menores valores indicam menor assimetria. Os autores verificaram o comportamento de seu índice nas quatro classes, estratificadas por idade, gênero, índice de educação e o volume total intracraniano. Para a análise estatística do índice, os autores utilizaram o teste MANCOVA seguido pelo *post-hoc* Tukey HSD (ABDI; WILLIAMS, 2010), ambos com nível de significância de 0,05, e para avaliar a lateralização da assimetria, os volumes esquerdo e direito dessas regiões foram comparados usando o teste *Student-t* (GOULDEN, 1939). Como resultado, os autores encontraram uma tendência geral de aumento da assimetria com o aumento da gravidade do diagnóstico (CN < sCCL < pCCL < DA). Além disso, foi encontrada uma tendência geral do volume esquerdo menor que o direito em oito das doze subestruturas hipocampais dos quatro grupos de diagnóstico.

Ardekani et al. (2019) analisaram a magnitude da assimetria da fração do parênquima² do hipocampo (HPF) utilizando 775 imagens de RM de 198 pacientes da base de dados OASIS (MARCUS et al., 2007), sendo 98 CN, 70 com CCL e 30 com DA. O cálculo do HPF foi realizado utilizando o módulo KAIBA do pacote

² O parênquima cerebral é constituído por substância cinzenta, onde estão localizados os corpos celulares dos neurônios, e por substância branca, onde se encontram os seus prolongamentos axonais que estabelecem a ligação entre diferentes áreas encefálicas e medulares.

ART³. A assimetria hipocampal foi calculada pelo valor absoluto da diferença entre os volumes dos HPF esquerdo e direito multiplicado por 100% e dividido pelo somatório dos dois volumes. Assim como Lucarelli et al. (2013), neste estudo, a magnitude da assimetria não considerou qual hipocampo apresentava maiores valores. Para as análises estatísticas os autores verificaram o comportamento de seu índice de assimetria nos gêneros feminino e masculino e o correlacionaram com a idade usando o teste ANOVA (SCHEFFE, 1999) seguido pelo teste *post-hoc* Tukey HSD (ABDI; WILLIAMS, 2010), ambos com nível de significância de 0,05. Como resultados os autores encontraram uma maior assimetria em homens que mulheres, aumento da assimetria conforme o aumento da idade, e aumento da assimetria conforme o aumento do declínio cognitivo (CN<CCL<DA). Conclusões que também são consistentes com os trabalhos de Shi et al. (2009) e Sarica et al. (2018).

A Tabela 3 apresenta um sumário dos principais trabalhos publicados na literatura que avaliaram a assimetria inter hemisférica estrutural e suas respectivas conclusões.

Tabela 3 – Resumo dos principais trabalhos publicados na literatura que analisam a assimetria estrutural.

Trabalho	Tipo	Volume	Grupos	Assimetria
Shi et al. (2009)	Meta-análise	Hipocampos	CN, CCL, DA	CN<CCL<DA, Volume Hip. Dir > Esq
Lucarelli et al. (2013)	Índice de assimetria	Hipocampos (Free Surfer + FIRST)	CN	F<M, Aumento assimetria com idade
Sarica et al. (2018)	Índice de assimetria	Hipocampos + subcampos (Free Surfer)	CN, sCCL, pCCL, DA	CN<sCCL<pCCL<DA, Volume Hip. Dir > Esq
Ardekani et al. (2019)	Índice de assimetria	Hipocampos (Free Surfer)	CN, CCL, DA	CN<CCL<DA, Aumento assimetria com idade

Os trabalhos descritos nesta subseção corroboram a existência da assimetria hipocampal na DA e no CCL, além de demonstrar, em Sarica et al. (2018), que os subcampos do hipocampo também possuem padrões distintos entre os volumes dos hipocampos esquerdo e direito. No entanto, embora esses trabalhos tenham analisado a assimetria em diferentes condições da DA, nenhum deles verificou seu poder preditivo.

³ <https://www.nitrc.org/projects/art>

Por esse motivo, este estudo apresenta uma nova metodologia para realizar a classificação de imagens de RM usando apenas atributos que apresentam diferenças significativas entre nossos grupos de diagnóstico (CN, CCL e DA leve) quando testados para assimetria. De maneira resumida, propusemos uma abordagem que utiliza a correlação entre estatísticas extraídas da resposta de filtros direcionais multi-escalas das regiões hipocampais esquerda e direita como atributos para a classificação. Essa abordagem utiliza informações direcionais entre os hemisférios cerebrais que são desconsideradas em análises puramente volumétricas. Além disso, a correlação identifica a natureza da associação entre os dois hemisférios, possibilitando a avaliação do padrão geral de assimetria que os pacientes de uma determinada classe possam apresentar, ao invés de comparar cada resposta direcional sozinha. Adicionalmente, propusemos a criação de um índice de assimetria e analisamos a distinção estatística entre os grupos de diagnóstico avaliados. A metodologia desenvolvida para a realização desta análise é apresentada no próximo capítulo.

3.3 Predição da idade biológica do cérebro

Este trabalho explorou também as associações entre a idade cronológica e a idade biológica estimada do cérebro de pacientes CN, com CCL e DA leve. O avanço da idade está associado a um aumento da prevalência de doenças cerebrais, principalmente neurodegenerativas, como a DA, doença de Parkinson e esclerose lateral amiotrófica (COLE; FRANKE, 2017). Portanto, avançar a compreensão do envelhecimento do cérebro e identificar biomarcadores para ajudar a melhorar o diagnóstico precoce dessas doenças é crucial. A estimação da idade do cérebro tem sido bastante estudada, pois pode identificar indivíduos em risco de idade avançada do cérebro e fornecer um biomarcador de problemas de saúde relacionados à idade, (COLE et al., 2017) dado que ter um cérebro saudável é considerado essencial para a saúde de maneira geral (COLE; FRANKE, 2017).

Os sistemas computacionais de auxílio ao diagnóstico usando imagens médicas estão em constante crescimento desde o surgimento das técnicas de aprendizado de máquina. A tarefa de predição da idade do cérebro foi proposta em trabalhos da literatura com o uso de diferentes técnicas, como a *Relevance Vector Machines*

(RVM) (FRANKE et al., 2010; FUJIMOTO et al., 2017; MADAN; KENSINGER, 2018), SVM (PARDAKHTI; SAJEDI, 2017; SU et al., 2011), *Gaussian process regression* (GPR) (COLE et al., 2017) e aprendizado profundo, principalmente com as CNNs (DINSDALE et al., 2021; COLE et al., 2017; PENG et al., 2021; ITO et al., 2018; HUANG et al., 2017; UEDA et al., 2019).

Os modelos de aprendizagem profunda, devido a terem alcançado os menores erros de predição da idade (PENG et al., 2021; BAUMGARTNER et al., 2019), têm se mostrado bastantes promissores (COLE et al., 2017). No entanto, as redes neurais profundas ainda enfrentam vários desafios em imagens médicas, por exemplo, pequena disponibilidade de dados rotulados para o treinamento, a necessidade de mais memória em GPUs para o processamento de dados 3D e a indisponibilidade de modelos 3D pré-treinados. Os pesquisadores tentam mitigar esses problemas usando algumas estratégias, que incluem a redução da resolução da imagem de entrada, o uso de pequenos *patches* (KAMNITSAS et al., 2017; LIU et al., 2018), recortes da imagem (DINSDALE et al., 2021; PARDAKHTI; SAJEDI, 2020) e fatias 2D (HUANG et al., 2017). No entanto, é difícil usar essas restrições sem perda de desempenho e/ou de informações relevantes.

Na subseção a seguir são descritos alguns estudos propostos para a tarefa de predição da idade do cérebro usando aprendizado profundo, visto que é a técnica mais utilizada na área e apresentou os menores erros de predição.

3.3.1 Trabalhos correlatos

Huang et al. (2017) treinaram um modelo 2D-CNN para a estimaco etria do crebro usando fatias especficas das imagens de RM. Os autores utilizaram 1099 imagens de indivduos CN com faixa etria entre 20 e 80 anos do centro de pesquisa *Aoba brain image* (SATO et al., 2003) e do projeto *Sendai Tsurugaya* (HOZAWA et al., 2004), sendo que 600 dessas imagens foram utilizadas para o treinamento e o restante para o teste da rede. A arquitetura da rede  baseada na VGG (SIMONYAN; ZISSERMAN, 2014), composta por cinco camadas convolucionais e de *pooling*. Nesse caso, cada camada convolucional  seguida por uma camada de ativao ReLU (NAIR; HINTON, 2010), sendo as ltimas trs camadas totalmente conectadas e a de sada uma funo de ativao linear. A entrada da rede consiste

em 15 fatias com resolução de 189×157 voxels. As fatias selecionadas correspondem às que contém a maior parte do encéfalo e, segundo os autores, o desempenho não foi prejudicado com a seleção. A predição etária final corresponde à mediana das predições individuais de cada fatia. Os hiperparâmetros utilizados consistem no decaimento da taxa de aprendizado (BENGIO, 2012) por época, com valor inicial de $10e - 4$, do algoritmo de otimização SGD (SUTSKEVER et al., 2013) com *momentum*, o MAE como função de perda, a regularização L2 (HANSON; PRATT, 1988) e três transformações afins aleatórias (rotações horizontais, recortes e translações) utilizadas como *data augmentation* (PEREZ; WANG, 2017). O código foi implementado usando o PyTorch (PASZKE et al., 2019) e o treinamento realizado em uma GPU NVidia TitanX. Como resultado, os autores obtiveram um MAE de 4 anos.

No mesmo ano, Cole et al. (2017) treinaram quatro modelos de 3D-CNNs para prever a idade do cérebro a partir de imagens de RM. Os autores utilizaram 2001 imagens de indivíduos com faixa etária entre 18 e 90 anos da base de dados *Brain-Age Healthy Control*⁴ (COLE et al., 2015), sendo que 1601 imagens foram utilizadas para o treinamento, 200 para validação e 200 para o teste. A arquitetura das redes é composta por cinco sequências de camadas, sendo elas: convolucionais, de ativação ReLU (NAIR; HINTON, 2010), *batch normalization* (IOFFE; SZEGEDY, 2015), *pooling* e a camada de saída linear. A metodologia do trabalho consiste no treinamento de duas redes que utilizam imagens dos tecidos GM e WM como entrada, e a combinação das duas redes, GM e WM, pré-treinadas, resultando no terceiro modelo. O quarto e último modelo consiste na utilização de imagens de entrada com o mínimo de pré-processamento possível para avaliar se a predição é sensível à tais etapas. Os hiperparâmetros utilizados foram os mesmos para todos os modelos, consistindo no uso do MAE como função de perda, o SGD com *momentum* (SUTSKEVER et al., 2013) como algoritmo de otimização, regularização L2 (HANSON; PRATT, 1988), *batch* de tamanho 28 e a utilização de decaimento da taxa de aprendizado (BENGIO, 2012) por época, com valor inicial de $10e-1$ e decaimento de 3%. Além disso, como estratégia de *data augmentation*

⁴ Imagens de 14 bases de dados públicas de RM: (*Brain-Age Healthy Control*): 184 ABIDE, 181 Beijing Normal University, 49 Berlin School of Brain & Mind, 12 CADDementia, 31 Cleveland Clinic, 322 ICBM, 561 IXI, 93 MCIC, 23 MIRIAD, 39 NEO2012, 160 NKI/Rockland sample, 288 OASIS, 24 WUSL, 36 TRAIN-39 (SAJEDI; PARDAKHTI, 2019).

(PEREZ; WANG, 2017) foram realizadas transformações aleatórias de translação e rotação. O código foi implementado usando o PyTorch e o treinamento realizado usando quatro GPUs NVidia TitanX. O melhor resultado foi obtido para o modelo treinado com imagens de GM, com MAE de 4,16 anos, coeficiente de correlação de Pearson (r) de 0,96 e RMSE de 5,31. Para o modelo sem pré-processamento foi obtido um MAE de 4,65 anos, r de 0,94 e RMSE de 6,46. Além disso, os autores aplicaram o método GPR para comparar os resultados. Nesse caso, os melhores valores foram obtidos da combinação de GM e WM, com MAE de 4,41 anos, r de 0,96 e RMSE de 5,43. Entretanto, para o GPR sem pré-processamento foi obtido um MAE de 11,81 anos, r de 0,57 e RMSE de 15,1.

Este trabalho criou um modelo bastante eficaz para a predição da idade do cérebro, inclusive com o mínimo de pré-processamento dos dados. Ao comparar os resultados com a técnica GPR, os autores obtiveram uma diferença considerável nos erros de predição, principalmente ao não utilizar o pré-processamento. No entanto, não foram relatadas análises do nível de degradação das imagens de entrada utilizadas para o experimento sem pré-processamento. Nesse caso, seria interessante quantizar os níveis de ruído e *bias field*, por exemplo, para verificar se outras imagens de entrada obteriam os mesmos resultados.

Ueda et al. (2019) treinaram um modelo de 3D-CNN para a estimação da idade biológica do cérebro de indivíduos CN a partir de imagens de RM. Os autores utilizaram 1101 imagens de indivíduos CN com faixa etária entre 20 e 80 anos providas do centro de pesquisa *Aoba brain image* (SATO et al., 2003) e do projeto *Sendai Tsurugaya* (HOZAWA et al., 2004), sendo que 768 dessas imagens foram utilizadas para o treinamento e o restante para o teste. A arquitetura da rede criada para esse trabalho consiste em quatro blocos, compostos por uma camada convolucional, seguida por uma de *batch normalization* (IOFFE; SZEGEDY, 2015), função de ativação ReLU (NAIR; HINTON, 2010) e uma de *pooling*. A seguir, são inseridas três camadas completamente conectadas e uma função de ativação linear como saída. As imagens, que possuem resolução de $189 \times 157 \times 156$ *voxels* após serem pré-processadas, são reduzidas para $95 \times 79 \times 78$ *voxels* com o objetivo de diminuir o custo computacional e o consumo de memória. Em relação aos hiperparâmetros, foi utilizado o algoritmo SGD com *momentum* (SUTSKEVER et al., 2013), o MAE como função de perda, regularização L2 (HANSON; PRATT, 1988), *batch*

de tamanho 16 e a utilização de decaimento da taxa de aprendizado (BENGIO, 2012) por época, com valor inicial de $5e^{-4}$ e decaimento de $1e^{-5}$. Como estratégia de *data augmentation* (PEREZ; WANG, 2017), foram aplicadas rotações no eixo x , recortes e translações aleatórias. O código foi implementado usando o TensorFlow (MARTÍN et al., 2015) e o Keras (CHOLLET et al., 2015). Como resultado, os autores obtiveram um MAE de 3,67 anos, r de 0,97 e RMSE de 4,71. Além disso, os autores compararam seus resultados com trabalhos que utilizaram PCA (FRANKE et al., 2010), atributos locais (ITO et al., 2018) e 2D-CNN (HUANG et al., 2017) e mostraram que seus resultados foram superiores.

Feng et al. (2020) treinaram um modelo de regressão 3D-CNN para a predição da idade do cérebro usando imagens de RM de indivíduos CN. Os autores utilizaram 2852 imagens de 2694 indivíduos únicos com faixa etária entre 18 e 97 anos de diferentes bases de dados, com o objetivo de obter uma representação mais diversificada da população saudável. A arquitetura da rede utilizada é bastante semelhante à VGG (SIMONYAN; ZISSERMAN, 2014), composta por camadas convolucionais intercaladas por camadas de *pooling* e um número crescente de atributos ao longo da profundidade. Outras particularidades incluem a conversão de todas as operações 2D para 3D e a troca da função de ativação *softmax* (GOODFELLOW; BENGIO; COURVILLE, 2016) por uma linear. Em relação aos hiperparâmetros, foi utilizado o algoritmo Adam (KINGMA; BA, 2014) para otimização usando o MAE como função de perda, *batches* de tamanho cinco, regularização L2 (HANSON; PRATT, 1988), taxa de aprendizado variando (BENGIO, 2012) entre $10e^{-4}$ e $10e^{-6}$ e *data augmentation* (PEREZ; WANG, 2017) com a utilização das imagens longitudinais presentes nas bases. O código foi implementado usando o TensorFlow (MARTÍN et al., 2015) e o Keras (CHOLLET et al., 2015). Como resultado, os autores obtiveram um MAE de 4,06 anos e r de 0,97. Os autores também avaliaram seu modelo usando uma base de dados diferente e obtiveram um MAE de 4,21 anos e r de 0,96. Além disso, foi realizada uma avaliação dos valores do MAE na presença de desbalanceamento dos dados. Nesse caso, observou-se um menor valor de MAE em faixas etárias com maior quantidade de amostras, o que poderia introduzir um viés nos resultados. Por esse motivo, os autores recomendaram o uso de técnicas que gerem uma distribuição uniforme dos dados ao longo das faixas etárias.

Pardakhti e Sajedi (2020) treinaram três modelos de regressão 3D-CNN para a predição da idade do cérebro usando imagens de RM de indivíduos CN com diferentes resoluções. Os autores utilizaram 562 imagens da base de dados IXI de indivíduos CN com faixa etária entre 20 e 86 anos, sendo que 400 imagens foram utilizadas para o treinamento e 162 para o teste. A arquitetura da rede foi composta por blocos cinco camadas de convolução e *pooling*, seguidas por uma camada de *batch normalization* (IOFFE; SZEGEDY, 2015) e uma camada totalmente conectada. A camada de saída é uma função de ativação linear. As imagens de entrada do primeiro modelo possuem resolução de $79 \times 95 \times 79$ *voxels*, já para o segundo e o terceiro modelo, os autores selecionaram fatias mais centrais cérebro, sendo a resolução de cada uma de $32 \times 95 \times 79$ *voxels* e $48 \times 95 \times 79$ *voxels*, respectivamente. O conjunto de hiperparâmetros consistiu da função de perda MSE, regularização L2 (HANSON; PRATT, 1988), *batch* de tamanho 8 e a taxa de aprendizado de 0,001. Os melhores resultados foram obtidos para o primeiro modelo, com um MAE de 5,149 anos e RMSE de 13,5. Os autores também avaliaram o desempenho dos modelos usando os vetores de atributos da rede que antecedem a última camada usando SVR e GPR e os resultados foram inferiores à 3D-CNN inicialmente proposta. Além disso, avaliaram a generalização do modelo usando 47 imagens de indivíduos CN e 22 com DA da base de dados da ADNI com idade superior a 65 anos. Para os indivíduos CN foi obtido um MAE de 6,64 e um RMSE de 5,19 e para DA, um MAE de 8,637 e RMSE de 9,94.

Ainda sobre o trabalho de Pardakhti e Sajedi (2020), os autores criaram modelos para a predição da idade do cérebro com diferentes resoluções e obtiveram valores de MAE um pouco acima do encontrado em trabalhos da literatura. Ao comparar seus resultados com outra base de dados, eles obtiveram uma diferença baixa entre as métricas, mostrando boa generalização de seu modelo. No entanto, não foram fornecidas informações sobre a distribuição das faixas etárias para a comparação com a DA, inviabilizando uma análise apropriada dos resultados obtidos pelo trabalho, visto que pode haver um viés causado pela idade mais avançada, ao invés do desenvolvimento da doença. Além disso, as métricas utilizadas apresentadas são insuficientes para se concluir se há uma super estimação etária para os cérebros com DA, dado que se baseiam em valores absolutos.

Peng et al. (2021) propuseram a criação de uma arquitetura de rede 3D-CNN

para prever a idade do cérebro a partir de imagens de RM. Os autores utilizaram 14503 imagens da base de dados *UK Biobank* (MILLER et al., 2016), sendo 12949 usadas para o treinamento, 518 para validação e 1036 para o teste. A arquitetura da rede foi baseada na VGG (SIMONYAN; ZISSERMAN, 2014) e contém cerca de três milhões de parâmetros. O modelo consiste em cinco blocos sequenciais compostos por uma camada convolucional, seguida por uma de *batch normalization* (IOFFE; SZEGEDY, 2015), *pooling* e função de ativação ReLU (NAIR; HINTON, 2010). A seguir, existe um bloco composto por uma camada convolucional, de *batch normalization*, e ReLU e por fim, camadas de *pooling*, *dropout* (SRIVASTAVA et al., 2014), totalmente conectada e *softmax* (GOODFELLOW; BENGIO; COURVILLE, 2016). A camada de saída, *softmax*, contém 40 neurônios, que representam a probabilidade da idade de predição de cada intervalo etário (dois em dois anos). A metodologia do trabalho consiste no treinamento de 20 modelos, sendo quatro tipos distintos de imagens de entrada, GM, WM e T1-*w* com correção linear e T1-*w* com correção não linear. Para cada entrada foram treinados cinco modelos com inicializações aleatórias dos parâmetros, em seguida os vinte resultados foram combinados (*ensemble*) e o resultado final corresponde à média de todos os modelos. Além disso, os autores aplicaram o algoritmo linear de *bias correction* (STEPHEN et al., 2019), para corrigir o viés presente nas predições. O código foi implementado usando o PyTorch (PASZKE et al., 2019) e o treinamento realizado em duas GPUs NVidia P100. O conjunto de hiperparâmetros utilizado consistiu da divergência de Kullback-Liebler, como função de perda, o SGD (SUTSKEVER et al., 2013) com *momentum* como algoritmo de otimização, regularização L2 (HANSON; PRATT, 1988), *batch* de tamanho 8 e a utilização da multiplicação por 0,3 da taxa de aprendizado (BENGIO, 2012) a cada 30 épocas, com valor inicial de 0,01. Como estratégia de *data augmentation* (PEREZ; WANG, 2017), foram aplicadas rotações no eixo sagital e translações aleatórias. O melhor resultado foi obtido para a combinação dos vinte modelos, com uma MAE de 2,14 anos. Além disso, os autores compararam seus resultados com versões 3D da ResNet18, ResNet50, ResNet101 e ResNet152 (HE et al., 2016) e a ElasticNet (JOLLANS et al., 2019), e os resultados do método proposto foram superiores, ou seja, com menor função de perda.

Dinsdale et al. (2021) desenvolveram duas arquiteturas 3D-CNNs para prever

a idade do cérebro para os gêneros feminino e masculino a partir de imagens de RM. Os autores usaram 12802 imagens do banco de dados *UK Biobank* (MILLER et al., 2016), sendo que para a CNN do gênero feminino foram utilizadas 6579 imagens para o treinamento, 732 para validação e 3654 para o teste. Para a CNN do gênero masculino foram utilizadas, respectivamente, 6223, 692 e 3231 imagens para o treinamento, validação e o teste. As imagens utilizadas possuíram resolução de $182 \times 218 \times 182$ voxels; no entanto, para reduzir a memória necessária, as dimensões x e y foram redimensionadas para 128×128 voxels, e apenas 20 fatias (de 182) foram usadas. As fatias foram escolhidas para incluir o máximo possível do cérebro e algumas informações do cerebelo. A arquitetura consistiu majoritariamente em uma adaptação da VGG (SIMONYAN; ZISSERMAN, 2014) para a regressão e operações 2D para 3D. O modelo final consistiu da combinação da predição de três redes treinadas com diferentes inicializações. O código foi implementado usando o Keras (CHOLLET et al., 2015) com o TensorFlow (MARTÍN et al., 2015) e o treinamento realizado em uma GPU NVidia Tesla P100. O conjunto de hiperparâmetros utilizado consistiu do MSE como função de perda, o RMSprop (TIELEMAN; HINTON, 2012) como algoritmo de otimização, um *batch* de tamanho 16 e utilização da divisão pela metade da taxa de aprendizado (BENGIO, 2012) a cada 15 épocas, com valor inicial de 10^{-2} . Os modelos foram treinados por no máximo 200 épocas, com o treinamento interrompido, caso o MSE de validação não melhorasse por 30 épocas. No total foram treinados seis modelos por gênero, sendo três usando imagens com correção linear e três com correção não linear. Os melhores resultados para o modelo de predição, tanto feminino quanto masculino, foram obtidos para as imagens com correção não-linear. No entanto, os autores mostraram que o uso dessas imagens para treinar CNNs pode fazer com que a rede seja movida por artefatos do processo de registro e remova indicadores sutis do envelhecimento, limitando a sua relevância clínica. Portanto, considerando os resultados com correção linear para o modelo feminino, foram obtidos os seguintes valores: MAE de 2,86, MSE de 13,12 e r de 0,87 e para o modelo masculino: MAE de 3,09, MSE de 15,13 e r de 0,86.

A Tabela 4 apresenta um sumário dos principais métodos encontrados na literatura para a predição da idade biológica do cérebro.

Os trabalhos descritos obtiveram bons desempenhos para a tarefa de predi-

ção da idade do cérebro, apresentando baixos valores de MAE em uma grande diversidade de bases de dados e quantidades de imagens. Em suma, a maioria utilizou versões modificadas da VGG (SIMONYAN; ZISSERMAN, 2014) com três alterações predominantes: *i*) adição de camadas de *batch normalization* (IOFFE; SZEGEDY, 2015), *ii*) conversão de todas as operações 2D para 3D e, por se tratar de um problema de regressão, *iii*) substituição da função de ativação *softmax* por uma linear (HUANG et al., 2017; JIANG et al., 2020; PENG et al., 2021; ITO et al., 2018). Além disso, o MAE foi a métrica mais utilizada como função de perda, assim como a regularização L2, transformações de *data augmentation*, a otimização usando SGD (WIJNHOVEN; WITH, 2010) com *momentum* e a estratégia de decaimento da taxa de aprendizagem. Por fim, devido à indisponibilidade de pesos pré-treinados em 3D, todos os modelos foram treinados do início ao fim; portanto, não foi utilizada transferência de aprendizado. Adicionalmente, nenhum dos trabalhos analisou as associações do erro médio de predição com o desenvolvimento de doenças cognitivas. Li et al. (2017) exploraram essas relações usando apenas os dois volumes hipocâmpais como atributos para um regressor e já obtiveram resultados significativos.

Tabela 4 – Resumo dos principais trabalhos publicados na literatura para a predição da idade biológica do cérebro.

	# Imagens [Intervalo de idade]	CNN	Tam batch	Função de perda	Otimização	Regularização	Taxa de aprendizado	Resultados		
Huang et al. (2017)	1099 [20-80]	2D	16	MAE	SGD + <i>Momentum</i>	L2 <i>Data augmentation:</i> translação + rotação + recortes	Decaimento por época	MAE: 4		
Cole et al. (2017)	2001 [18-90]	3D	28	MAE	SGD + <i>Momentum</i>	L2 <i>Data augmentation:</i> translação + rotação	Decaimento por época	MAE: 4,16 RMSE: 5,31 r: 0,96		
Ueda et al. (2019)	1101 [20-80]	3D	16	MAE	SGD + <i>Momentum</i>	L2 <i>Data augmentation:</i> translação + rotação + recortes	Decaimento por época	MAE: 3,67 RMSE: 4,71 r: 0,97		
Feng et al. (2020)	2694 [18-97]	3D	5	MAE	Adam	L2 <i>Data augmentation:</i> longitudinal	Intervalo de variação	MAE: 4,21 r: 0,96		
Pardakhti e Sajedi (2020)	562 [20-86]	3D	8	MSE	SGD + <i>Momentum</i>	L2	Valor fixo	MAE: 5,149 RMSE: 13,5		
Peng et al. (2021)	14503 [44-80]	3D	8	Kullback-Liebler	SGD + <i>Momentum</i>	L2 <i>Dropout</i> <i>Data augmentation:</i> translação + rotação (plano sagital)	Decaimento a cada 30 épocas	MAE: 2,14		
Dinsdale et al. (2021)	12802 [44-80]	3D	16	MSE	RMSprop	<i>Early stopping</i>	Decaimento a cada 30 épocas	MAE	Fem	Masc
								RMSE	2,86	3,09
								r	13,12	15,13
									0,87	0,86

Com base no que foi exposto, um dos objetivos desta pesquisa foi explorar as associações do erro médio de predição da idade do cérebro de pacientes da mesma idade do grupo CN com pacientes dos grupos CCL e DA leve. Para isso, combinamos as melhorias recentes em arquiteturas CNNs e projetamos dois experimentos para a predição da idade biológica do cérebro usando apenas as regiões do hipocampo. A análise regional pode ser mais interessante, visto que há críticas na literatura sobre condensar as informações do cérebro inteiro em um único número (COLE; FRANKE, 2017). Além disso, as diferenças entre as intensidades da atrofia hipocampal no envelhecimento normal e durante o desenvolvimento da DA é conhecida, como a existência da diminuição do volume de GM durante a idade adulta (GOOD et al., 2001; RESNICK et al., 2003), que ocorre de maneira mais acentuada durante o desenvolvimento da doença (GUADALUPE et al., 2014; WOLARD; HECKERS, 2012). A metodologia desenvolvida para a realização desta análise é apresentada no próximo capítulo.

3.4 Considerações finais

Este capítulo apresentou os principais trabalhos presentes na literatura que possuem intersecções técnicas e teóricas com esta pesquisa. Os trabalhos foram divididos de acordo com as três categorias abordadas neste trabalho, sendo: atrofias estruturais, assimetrias estruturais e o envelhecimento biológico do cérebro.

Para os trabalhos de atrofias estruturais, foram descritas com maior detalhamento técnicas que utilizam *landmarks*, ROIs, e seleção de atributos e objetivam a classificação da DA. Em seguida, discutimos as limitações dos métodos e, considerando-as, propusemos a primeira abordagem desenvolvida nesta pesquisa.

Para os trabalhos de assimetrias estruturais, foram descritas as metodologias utilizadas para a realização de análises estatísticas da assimetria cerebral com principal enfoque na DA. Nesse caso, observamos uma limitação em comum, que é a ausência do uso dos atributos extraídos para a classificação. Sendo assim, propusemos a segunda abordagem desenvolvida nesta pesquisa.

Para os trabalhos sobre a idade biológica do cérebro, foram descritas as metodologias de treinamento das redes CNN utilizadas, com informações sobre os hiperparâmetros e estratégias de convergência. Nesse caso, verificamos que tal

tarefa de predição tem sido bastante explorada, com erros de predição pequenos, e motivada pela diferença que pode existir quando comparada com doenças cognitivas, como o desenvolvimento da DA. No entanto, essa comparação tem sido pouco explorada, sendo esta a motivação da terceira abordagem desenvolvida nesta pesquisa.

A metodologia de cada abordagem é descrita no capítulo seguinte.

Capítulo 4

Materiais e Métodos

Este capítulo apresenta a descrição das bases de dados e dos materiais e métodos utilizados nesse trabalho, englobando suas técnicas, características e como foram aplicadas em cada uma das abordagens desenvolvidas.

4.1 Visão geral

Conforme descrito no Capítulo 1, o principal objetivo desta pesquisa foi estudar e desenvolver técnicas computacionais automáticas capazes de classificar alterações estruturais cerebrais em imagens de RM para auxiliar o diagnóstico de pacientes com CCL e DA leve. Para alcançar tal objetivo, propusemos três abordagens que buscaram analisar conceitos distintos e complementares da DA, a nomear a atrofia, a assimetria, e a análise do erro de estimação da idade biológica do cérebro. Além disso, propusemos unir as previsões das três abordagens para verificar a complementaridade dos resultados obtidos por cada uma delas. A Figura 6 apresenta um fluxograma que indica a sequência de etapas utilizadas no desenvolvimento nesta pesquisa. Cada etapa do fluxograma é detalhada nas seções que seguem.

A literatura tem trazido o uso de redes profundas para o auxílio ao diagnóstico da DA, no entanto, devido à baixa quantidade de imagens disponíveis de pacien-

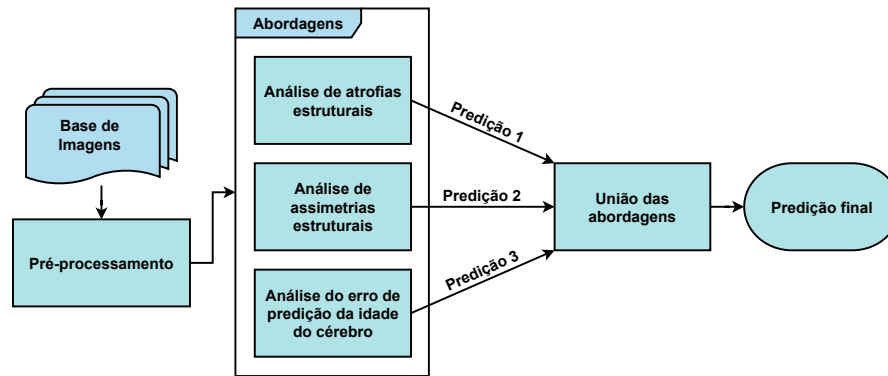


Figura 6 – Visão geral das principais abordagens desenvolvidas nesta pesquisa.

tes com declínio cognitivo, ainda existe uma grande dificuldade no treinamento dessas redes. Por esse motivo, as abordagens desenvolvidas nesta Tese exploram metodologias que utilizam engenharia de atributos quando os modelos são treinados em classes com declínio cognitivo, como nas abordagens de assimetria e atrofia, e aprendizado profundo na abordagem de estimação da idade biológica do cérebro, visto que o modelo é treinado utilizando somente imagens de pacientes cognitivamente saudáveis.

4.2 Base de imagens

Para a realização desta pesquisa foram utilizadas as bases de imagens 3D de RM T1-*w* do *Neuroimage Analysis Center* (NAC) (HALLE et al., 2017), da *Alzheimer’s Disease Neuroimaging Initiative* (ADNI) (JACK et al., 2017) e da *Information eXtraction from Images* (IXI¹).

4.2.1 NAC Brain Atlas

O NAC é um centro de pesquisa afiliado ao *Surgical Planning Laboratory* e à Universidade de *Harvard*. Sua base de dados contém 149 malhas triangulares 3D, cada uma representando uma estrutura cerebral distinta, rotuladas por um especialista. As malhas estão espacialmente alinhadas às imagens em T1-*w* e T2-*w* de um paciente do sexo masculino, saudável, de 42 anos. As imagens foram

¹ <<https://brain-development.org/ixi-dataset/>>

adquiridas usando um *scanner* Siemens de 3T com um tamanho de *voxel* isotrópico de 0,75 mm. Essa base de dados também contém uma imagem redimensionada de ambas as aquisições, T1-*w* e T2-*w*, com resolução isotrópica de 1 mm, e tamanho de $256 \times 256 \times 256$ *voxels*. Para este estudo, apenas a imagem T1-*w* redimensionada e as malhas triangulares hipocampais esquerda e direita 3D foram utilizadas.

4.2.2 ADNI

A iniciativa ADNI foi lançada em 2003 por um grupo de instituições de pesquisa, empresas farmacêuticas privadas e organizações não-governamentais, com a ajuda de pesquisadores de todo o mundo. A iniciativa tem buscado por maneiras de determinar a progressão da DA, desenvolvendo biomarcadores clínicos para a detecção da doença em estágios iniciais. O projeto está dividido em três fases, ADNI1, ADNI GO e ADNI2, as quais podem ser acessadas pelo *USC's Laboratory of Neuro Imaging's* (LONI) na página de *Image & Data Archive* (IDA)². Sua base de dados contém imagens de RM estruturais, funcionais (fRM), *Positron Emission Tomography* (PET), *Diffusion Tensor Imaging* (DTI) e informações genéticas e bioquímicas.

Para este estudo foram utilizadas imagens de RM adquiridas pela sequência *Magnetization Prepared Rapid Gradient Echo* (MPRAGE³) e *scanners* de 1.5T e 3T de três principais fabricantes (Philips, General Electric e Siemens). Elas foram adquiridas em mais de 50 locais nos EUA e no Canadá⁴ e totalizam 762 imagens de indivíduos com faixa etária entre 70 e 85 anos, divididos em três grupos de diagnóstico: CN, CCL e DA leve. A Tabela 5 apresenta os dados demográficos dos indivíduos.

4.2.3 IXI

A base de dados IXI faz parte do projeto *Brain Development* do *Imperial College London*. Sua base contém cerca de 600 imagens de RM estruturais de indi-

² <<https://ida.loni.usc.edu/login.jsp?project=ADNI>>

³ “*These MPRAGE files are considered the best in the quality ratings and have undergone gradwarping, intensity correction, and have been scaled for gradient drift using the phantom data*”. - <<http://adni.loni.usc.edu/methods/mri-tool/mri-analysis/>>

⁴ <<http://adni.loni.usc.edu/about/centers-cores/study-sites/>>

Tabela 5 – Conjunto de imagens da base de dados ADNI.

	CN	CCL	DA
No. pacientes	302	251	209
Idade	75.79 ± 4.14	76.80 ± 4.30	77.12 ± 4.43
MMSE	29.56 ± 0.50	27.04 ± 0.82	22.15 ± 2.76
Gênero (F:M)	151:151	80:171	94:115

MMSE: *Mini-Mental State Examination*.

vídus CN, com ponderação T1-*w*, T2-*w* e *proton density* (PD), além de imagens *Magnetic resonance angiography* (MRA) e DTI em 15 direções. As imagens foram coletadas em três diferentes hospitais, o *Hammersmith Hospital* (HH), o *Guy's Hospital* (GH) e o *Institute of Psychiatry* (IOP), que utilizam, respectivamente, os *scanners* 3T e 1.5T da Philips e 1.5T da *General Electric* (GE). Para este estudo foram utilizadas as imagens de RM dos três hospitais com ponderação em T1-*w* com faixa etária entre 20 e 80 anos, totalizando 557 imagens. A Tabela 6 apresenta os dados demográficos dos indivíduos.

Tabela 6 – Conjunto de imagens da base de dados IXI.

	GH	HH	IOP
No. pacientes	310	180	67
Idade	50.19 ± 15.4	47.17 ± 16.55	41.72 ± 15.81
Gênero (F:M)	350:135	186:87	86:24

4.3 Pré-processamento

As técnicas de pré-processamento apresentadas no Capítulo 2 foram utilizadas em todas as abordagens desta pesquisa. Primeiramente, todas as imagens de estudo apresentadas pelas Tabelas 5 e 6 foram processadas para a redução de ruído, correção do *bias field*, padronização de intensidades e correção afim, sendo que

às duas últimas etapas foram realizadas usando a imagem T1- w da NAC como referência. Em seguida, foi realizada a extração da região do encéfalo e, para delimitar adequadamente as regiões de interesse, foram realizados corregristos deformáveis entre a imagem da NAC e todas as imagens de estudo. Nesse caso, cada imagem de estudo foi utilizada como referência apenas para obter a transformação deformável, a qual foi aplicada aos modelos das malhas hipocampais 3D da NAC para o seu posicionamento adequado.

4.3.1 Parâmetros e análises qualitativas

Para a redução de ruído foi utilizado o algoritmo NLM (subseção 2.3.1) com os parâmetros padrões definidos pelos autores, presentes em (COUPÉ et al., 2008). Nesse caso, o valor do desvio-padrão (σ) do ruído foi estimado automaticamente, sendo utilizada uma vizinhança de 26 voxels definida por $(2d + 1)^3$, com $d \approx 1$ e volume de 1331 voxels, definido por $(2M + 1)^3$, com $M = 5$. A Figura 7 ilustra um exemplo da aplicação do NLM a uma imagem de RM da base de dados ADNI. É possível notar que o algoritmo conseguiu reduzir o ruído sem causar perdas significativas de informações estruturais na imagem, fato evidenciado principalmente na região do tronco cerebral, logo abaixo do corpo caloso, indicado pela seta branca das imagens sagitais.

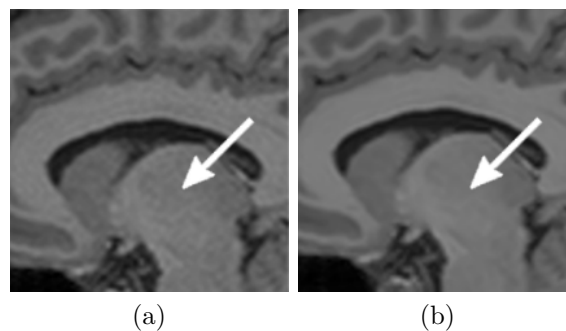


Figura 7 – Vista sagital de uma imagem de RM com ponderação em T1- w (a) antes e (b) após aplicação do NLM.

Para a correção do *bias field* foi utilizado o algoritmo N4-ITK (TUSTISON et al., 2010) (subseção 2.3.2) com os parâmetros definidos pelos autores no artigo

(TUSTISON et al., 2010). Nesse caso, o número de iterações foi de $500 \times 400 \times 300$, o valor de largura total na metade máxima foi de 0,15, o limiar de convergência de 0,001 e, para diminuir o tempo de execução, um fator de redução da imagem (*downsampling*) de 4. A Figura 8 ilustra um exemplo da aplicação do N4-ITK a uma imagem de RM da base de dados IXI. É possível verificar visualmente a correção, especialmente na área inferior esquerda das imagens sagitais.

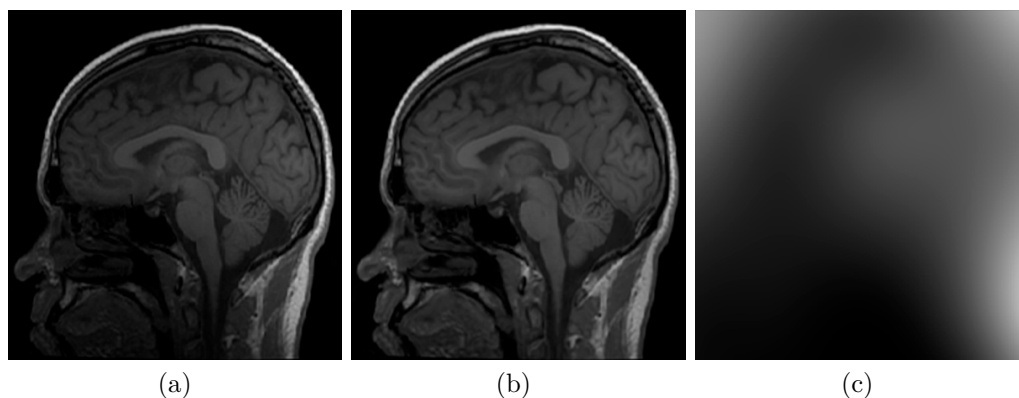


Figura 8 – Vista sagital de uma imagem de RM com ponderação em T1-*w* (a) antes, (b) após a aplicação do N4-ITK e (c) *bias field* removido.

Para a padronização de histogramas foi utilizada uma variação do algoritmo proposto por Nyúl, Udupa e Zhang (2000) (subseção 2.3.3). Nesse caso, a imagem T1-*w* da NAC foi utilizada como referência e as móveis foram as de estudo (Tabelas 5 e 6). Os parâmetros incluem o uso de 128 bins na quantização dos histogramas e três casamentos entre seus bins. A Figura 9 mostra um exemplo da padronização dos histogramas, exibindo em a) o histograma da imagem de referência; em b) o histograma de imagem clínica antes e em c) após a aplicação do algoritmo. É possível verificar visualmente que, após a aplicação do método, o histograma da imagem clínica se assemelha mais ao histograma da imagem de referência.

Para o corregristo afim foi utilizado o algoritmo proposto por (OURSELIN et al., 2001) e (OURSELIN; STEFANESCU; PENNEC, 2002) (subseção 2.3.4), em que a imagem da base de dados NAC foi utilizada como referência e as imagens de estudo foram as imagens móveis. A técnica foi utilizada com os principais valores de parâmetros disponíveis em http://cmictig.cs.ucl.ac.uk/wiki/index.php/Reg_aladin e com os valores de $-pv$ e $-pi$ iguais a 25. A Figura 10 mostra um

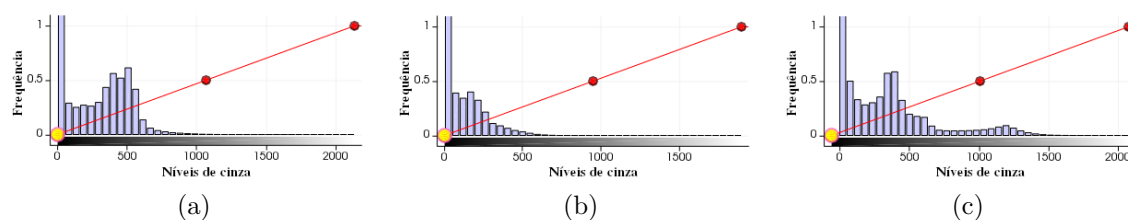


Figura 9 – (a) Histograma da imagem de referência; (b) histograma de uma imagem de estudo antes e (c) após a aplicação do algoritmo.

exemplo da aplicação do corregristo afim. Neste caso, uma imagem de estudo da ADNI e a imagem da NAC foram sobrepostas utilizando o padrão “tabuleiro de xadrez” que secciona ambas as imagens em blocos e monta uma representação final combinando os blocos subsequentes vindos alternadamente das duas imagens. Nota-se que a transformação afim conseguiu alinhar as estruturas cerebrais, fato evidenciado principalmente pela continuidade do corpo caloso dos blocos da região dessa estrutura, indicado pela seta branca da imagem sagital da transformação afim.

Para o corregristo deformável, aplicado somente para o posicionar as malhas dos hipocampos, foi utilizado o algoritmo proposto por (MODAT et al., 2010; RUECKERT et al., 1999) (subseção 2.3.4). Nesse caso, as imagens de estudo foram utilizadas como referência e a imagem da NAC foi a móvel de modo a obter a transformação deformável que mapeia a imagem móvel na imagem fixa. A transformação foi aplicada às malhas hipocampais da NAC para melhor posicioná-las nas imagens de estudo. Os valores dos parâmetros utilizados estão disponíveis em http://cmictig.cs.ucl.ac.uk/wiki/index.php/Reg_f3d.

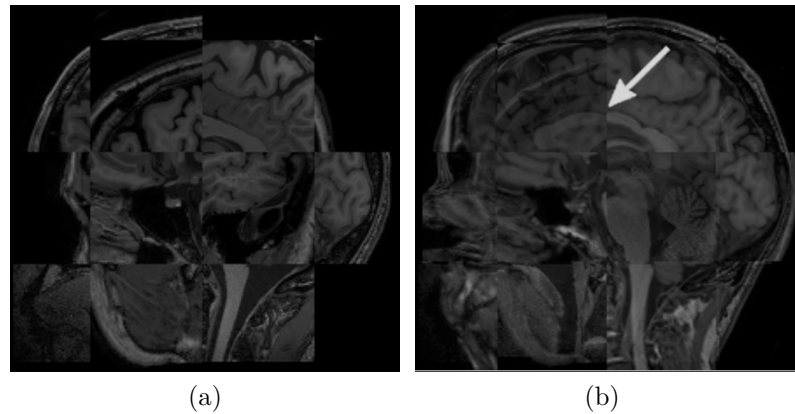


Figura 10 – Comparação na vista sagital entre o “tabuleiro de xadrez” formado por uma imagem de estudo e a imagem da NAC (a) antes e (b) após o corregristo afim.

Para a extração da máscara do encéfalo foi utilizada a técnica ROBEX (IGLESIAS et al., 2011) (subseção 2.3.5) com os parâmetros definidos em (IGLESIAS et al., 2011). A Figura 11 mostra um exemplo da aplicação do ROBEX em uma imagem de RM. Esta etapa foi utilizada apenas para as abordagens: análise de atrofia estruturais e análise de assimetrias estruturais.

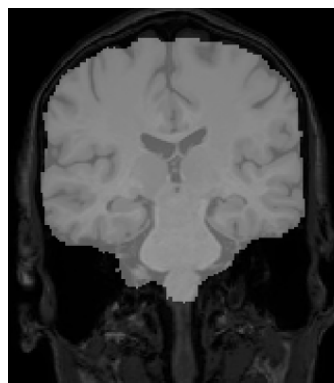


Figura 11 – Vista coronal de uma imagem de RM com a máscara do encéfalo obtida pelo ROBEX sobreposta à imagem original.

4.4 Análise de atrofas estruturais

Esta abordagem consiste em analisar atrofas estruturais a partir da seleção e posterior análise dos descritores de pontos salientes discriminativos entre duas populações nas regiões dos hipocampos. A Figura 12 contém um fluxograma que indica a seqüência de etapas utilizadas no desenvolvimento desta análise.

As imagens da base de dados ADNI foram inicialmente pré-processadas e os pontos salientes foram detectados e os seus respectivos descritores obtidos. Em seguida foi criado um atlas probabilístico de pontos salientes usando apenas as imagens de pacientes CN da base de dados IXI. O principal objetivo do atlas é determinar os pontos salientes mais estáveis entre o conjunto de imagens de pacientes CN, ou seja, pontos que são mais frequentemente detectados numa mesma região do espaço 3D das imagens CN, e assim utilizá-los para análises futuras.

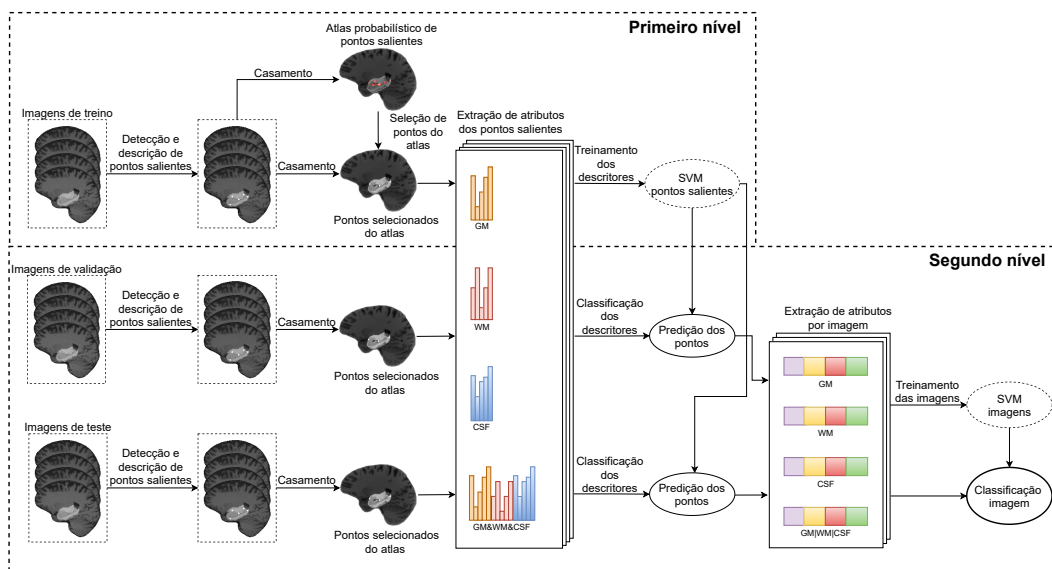


Figura 12 – Fluxograma da seqüência de etapas utilizadas no desenvolvimento da análise de atrofas estruturais.

As análises subsequentes foram realizadas em dois níveis de classificação: o nível de pontos salientes e o nível da imagem. No primeiro nível, pontos estáveis de um atlas probabilístico de pontos salientes (POLONI et al., 2018), mostrando diferenças estatisticamente significativas entre pares de populações diagnósticas (CN×CCL, CCL×DA e CN×DA), foram selecionados usando um conjunto de

imagens de treinamento. Em seguida, os descritores (atributos) de todos os pontos do conjunto de imagens de treinamento que se casaram com os pontos selecionados do atlas foram usados para treinar classificadores SVM. No segundo nível, usando um conjunto de imagens de validação, os pontos selecionados do atlas e os SVMs já treinados, foi realizada uma avaliação quantitativa das classificações dos SVMs (do nível de pontos salientes) para gerar os atributos do segundo nível. Por fim, usando os novos atributos, classificadores SVMs foram treinados para o nível da imagem e o conjunto de imagens de teste foi classificado usando todos os modelos treinados.

4.4.1 Definição das regiões de interesse

Esta etapa consiste na definição e modificação das máscaras binárias utilizadas para restringir as ROIs utilizadas nesta abordagem. A partir das duas máscaras binárias dos hipocampus direito e esquerdo posicionadas pelo pré-processamento, foi aplicado o operador morfológico de dilatação com um elemento estruturante definido por uma esfera 3D de raio de 8 *voxels* visando aumentar a área de análise e garantir que possíveis variações das estruturas representadas sejam levadas em consideração.

4.4.2 Seleção de pontos salientes

De modo a capturar mudanças cerebrais estruturais diferenciais entre cada uma das populações diagnósticas analisadas nesta pesquisa, ou seja, CN×CCL, CCL×DA, e CN×DA, foram utilizados os descritores dos pontos salientes para realizar uma comparação entre os pontos salientes do atlas e das imagens de treinamento. Os pontos foram detectados nas regiões hipocampais de cada imagem de treinamento usando a técnica de PC e descritos usando o descritor local (*spider-web*) mencionado no Capítulo 2, subseção 2.7.2. Em seguida, para localizar as posições correspondentes, foi realizado o casamento entre os pontos salientes de cada imagem de treinamento e os do atlas. Então, foi executado o *Welch's t-test* (WELCH, 1947) para comparar os grupos de diagnóstico e gerar um mapa de p-valor. Nesse caso, a distância χ^2 (resultante do casamento) foi utilizada para calcular o teste estatístico pareado das populações. Finalmente, os pontos do atlas

foram selecionados, caso fossem menores que um nível de significância definido (por exemplo, $\alpha = 0,01$ ou $\alpha = 0,05$). A distribuição dos dados foi testada para a normalidade e passou em todos os experimentos. O fluxograma das etapas para a seleção de pontos salientes é mostrado na Figura 13.

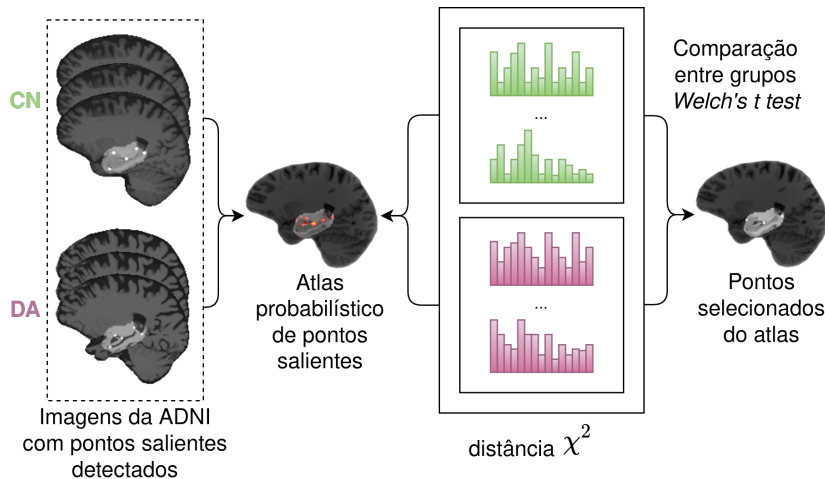


Figura 13 – Fluxograma para a seleção dos pontos salientes do atlas.

4.4.3 Extração de atributos dos pontos salientes

Devido aos pontos salientes selecionados serem provindos de mapas de p-valor criados da comparação de grupos de diagnóstico, eles estão diretamente associados às mudanças cerebrais estruturais diferenciais entre essas duas populações, fazendo com que tais pontos sejam (provavelmente) mais discriminativos para a classe de sua própria imagem. Portanto, foram extraídos atributos ao redor de cada ponto mencionado anteriormente e seus descritores (atributos) foram treinados usando a classe da imagem correspondente como o rótulo.

Extração

Os atributos foram extraídos individualmente de mapas de imagens probabilísticas das imagens de RM dos tecidos WM, GM e CSF. Especificamente, as imagens do cérebro, após serem pré-processadas para a extração do encéfalo pela técnica ROBEX (IGLESIAS et al., 2011) (seção 4.3), foram usadas como entrada para

o software FAST (versão 4.1) (ZHANG; BRADY; SMITH, 2001). O software foi executado para realizar a segmentação da imagem nesses três tipos diferentes de tecidos mencionados e gerar os mapas probabilísticos dos tecidos cerebrais GM, WM e CSF. Dessa maneira, ao redor de cada ponto saliente casado com os pontos selecionados do atlas, foi computado o histograma log-esférico global proposto por Villa-Pinto e Ferrari (2016) (subseção 2.7.2), usando as imagens probabilísticas dos três tecidos como fonte de informação. A Figura 14 mostra as imagens utilizadas para a extração dos atributos.

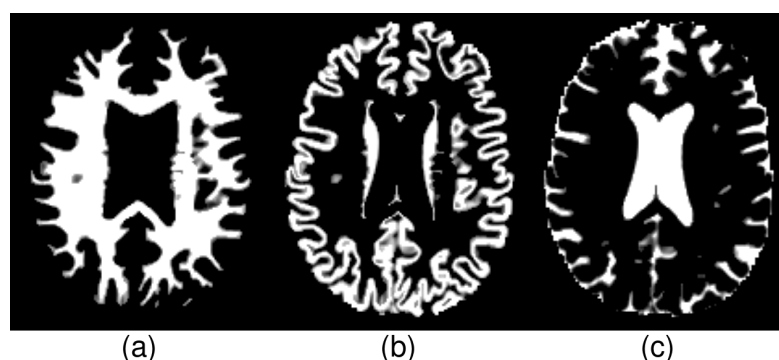


Figura 14 – Mapas probabilísticos das imagens de RM utilizados como fonte de informação para a extração dos atributos: (a) WM, (b) GM, e (c) CSF.

Organização

Usando os atributos extraídos de cada um dos três tecidos cerebrais e sua concatenação (WM|GM|CSF), quatro classificadores SVM foram treinados usando as imagens de treinamento. Nesse caso, cada descritor foi considerado uma amostra individual, portanto, não relacionado aos demais descritores extraídos de uma mesma imagem.

4.4.4 Classificação das imagens

No nível da imagem, após a etapa de seleção de pontos salientes, espera-se que os pontos tenham sido corretamente classificados em sua classe correspondente e seus atributos contribuam para a caracterização da evolução do Alzheimer. Além

disso, devido à classificação ser realizada em dois níveis, mesmo que existam pontos salientes classificados erroneamente no primeiro nível, espera-se que o impacto seja mínimo ao atingir o nível da imagem.

O vetor de atributos utilizado neste nível foi formulado com base em uma análise quantitativa da saída dos classificadores do primeiro nível. Dessa maneira, para cada região hipocampal, o vetor de atributos incluiu a quantidade de vezes que cada rótulo de classe foi predito ($\#classe_1, \#classe_2$), a sua probabilidade acumulada ($\sum \text{prob}_{classe\ 1}, \sum \text{prob}_{classe\ 2}$) e o total de pontos. Portanto, o tamanho do vetor resultante é de dez, ou seja, cinco atributos por região, esquerda e direita.

A seguir é descrita a metodologia utilizada na análise de assimetrias estruturais. As configurações experimentais, resultados e discussões desta análise são apresentados no capítulo seguinte.

4.5 Análise de assimetrias estruturais

Esta abordagem consiste em analisar as assimetrias estruturais hipocampais que são decorrentes do desenvolvimento do Alzheimer, desde o comprometimento cognitivo leve até o desenvolvimento mais avançado da doença. A Figura 15 contém um fluxograma que indica a sequência de etapas utilizadas no desenvolvimento desta abordagem.

As imagens da base de dados ADNI foram inicialmente preprocessadas e seus dois hemisférios cerebrais, direito e esquerdo rotacionado, extraídos. Em seguida, as imagens de cada hemisfério cerebral foram processadas via operação de convolução com um banco de filtros log-Gabor 3D, especialmente projetado para a obtenção das imagens direcionais. Logo após, para a utilização de seus três autovalores, a matriz de momentos de segunda ordem de inércia foi computada. Depois disso, restringindo a ROI ao hipocampo, com uma máscara binária, foram calculados os quatro primeiros momentos estatísticos (Média, Variância, *Skewness* e *Kurtosis*) e a correlação cruzada normalizada entre os vetores, esquerdo rotacionado e direito, de cada momento estatístico. Então os atributos resultantes foram utilizados para treinar classificadores SVM e classificar as imagens. Por fim, a partir da correlação cruzada, um índice de assimetria (*IA*) foi calculado e testado

entre os diferentes grupos de diagnóstico e gêneros utilizando o teste estatístico ANOVA.

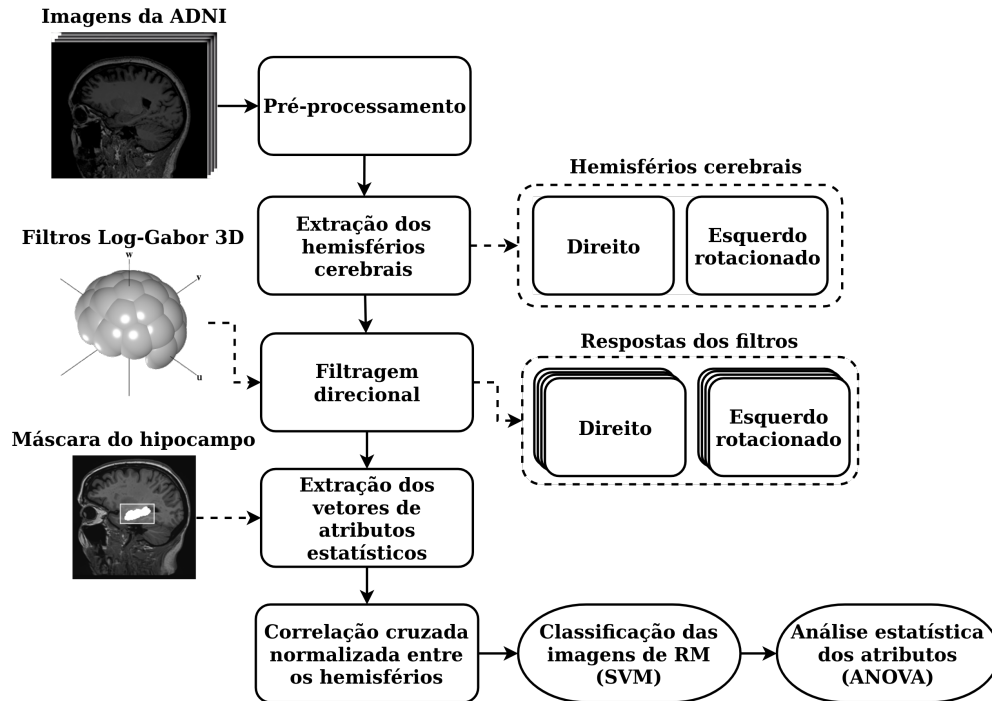


Figura 15 – Fluxograma da sequência de etapas utilizadas no desenvolvimento da análise de assimetrias estruturais.

4.5.1 Plano sagital médio

A identificação correta do plano que divide os hemisférios cerebrais é uma tarefa crucial para uma avaliação correta de assimetrias estruturais inter-hemisféricas. Conforme o estudo realizado por Liu (2009), muitos dispositivos produzem imagens cerebrais desorientadas, ou seja, com inclinações e distorções (LIU; COLLINS; ROTHFUS, 2001), que podem induzir um erro de inspeção visual e, muitas vezes, produzir uma interpretação clínica falsa em trabalhos que analisam a assimetria cerebral. Isso pode ser causado devido à imobilidade dos pacientes, inexperiência dos técnicos e imprecisão dos sistemas de calibração (LIU et al., 2006). Tal problema é corrigido com a detecção e ajuste do MSP, visto que são analisadas as

relações inter-hemisféricas, possibilitando, assim, a realização de avaliações mais precisas dos dados por um especialista ou pelo computador.

Nesta pesquisa, o MSP foi identificado usando o algoritmo proposto por Ferrari, Villa-Pinto e Moreira (2016), mencionado na subseção 2.7.4. Após a identificação do plano, o cérebro foi transladado para seu centro de gravidade e rotacionado, de maneira que o MSP ficasse paralelo às coordenadas sagitais. A seguir, o hemisfério esquerdo foi rotacionado em 180° no eixo sagital com o propósito de obter os dois hemisférios aproximadamente na mesma localização espacial. A Figura 16 mostra um exemplo da detecção do MSP, exibindo em (a) a detecção do MSP, (b) o hemisfério direito e em (c) o hemisfério esquerdo rotacionado.

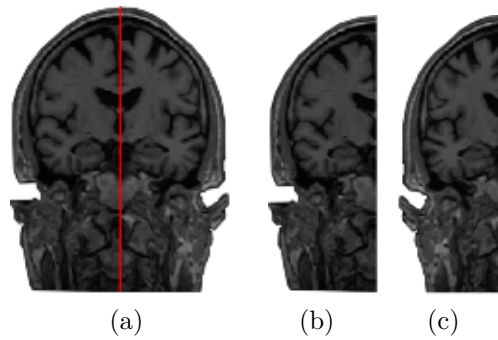


Figura 16 – Vista coronal de uma imagem de RM com detecção do MSP, em (a) um exemplo da detecção, (b) o hemisfério direito e (c) o hemisfério esquerdo rotacionado.

4.5.2 Filtragem direcional

Conforme descrito na seção 2.6 do Capítulo 2, as imagens obtidas do processamento da filtragem direcional são resultantes da combinação das respostas obtidas da convolução de uma imagem com os filtros log-Gabor 3D para diferentes escalas e orientações, em conjunto com as imagens correspondentes aos três autovalores (λ_0 , λ_1 e λ_2) da matriz $M_{DR}(\mathbf{x})$. Para evitar artefatos causados por bordas na imagem filtrada resultante, o processo de filtragem foi aplicado a toda a região do encéfalo de cada hemisfério do cérebro. Após a filtragem, a máscara binária do hipocampo foi utilizada para restringir a região de interesse e permitir a extração dos atributos assimétricos.

4.5.3 Análise entre hemisférios

A primeira etapa da análise entre os hemisférios consiste em uma modificação nas máscaras binárias hipocampais obtidas no pré-processamento. Primeiramente, a máscara referente ao hemisfério esquerdo foi rotacionada em 180° . Em seguida, às duas máscaras dos hipocampos (direita e esquerda rotacionada) foram combinadas, com o operador de união, para gerar uma máscara única que foi então utilizada para as análises dos hemisférios. Por fim, a máscara resultante foi expandida com a operação morfológica de dilatação usando uma esfera 3D de raio de 2 voxels como elemento estruturante. A expansão da máscara foi realizada para garantir a inclusão dos hipocampos, levando-se em consideração a variabilidade natural dessa estrutura na população.

Extração de atributos assimétricos

A etapa de filtragem gerou um total de 26 imagens, ou seja, $23 DR_{\Theta} + 3$ mapas de autovalores, λ_0, λ_1 , e λ_2 , respectivamente. Com o propósito de caracterizar a distribuição das respostas dos filtros direcionais em cada hemisfério, os quatro primeiros momentos estatísticos (Média, Variância, *Skewness* e *Kurtosis*) foram calculados a partir das intensidades dos voxels que estão dentro da máscara do hipocampo, resultando em quatro vetores de atributos (de tamanho $N = 26$, cada) por hemisfério.

Além da extração dos atributos assimétricos hipocampais, diagramas de rosas em 3D (MARDIA; JUPP, 2009) residuais foram criados de modo a ilustrar a assimetria hemisférica hipocampal e as diferenças entre as respostas direcionais das estatísticas extraídas a partir das 23 imagens direcionais (DR_{Θ}). Os diagramas foram criados a partir da subtração absoluta dos diagramas dos hipocampos, esquerdo (rotacionado) e direito, utilizando a mesma resolução angular, ângulos de azimute e elevação utilizados para gerar os filtros log-Gabor 3D utilizados na filtragem. Dessa maneira, cada cone no diagrama mapeia a magnitude residual de uma dada medida estatística calculada a partir da região hipocampal de uma imagem filtrada.

A Figura 17 ilustra as vistas coronais das regiões dos hipocampos direito e esquerdo rotacionado e os correspondentes diagramas de rosas residuais 3D criados

para a medida de Variância obtida para as imagens de RM de três pacientes pertencentes aos grupos CN, CCL e DA. Na Figura é possível visualizar a extensão das assimetrias hemisféricas nos grupos de diagnóstico, visível tanto nas imagens hipocampais quanto nos diagramas residuais. É possível verificar também, uma diminuição dos hipocampus e um aumento nos resíduos conforme o aumento do declínio cognitivo.

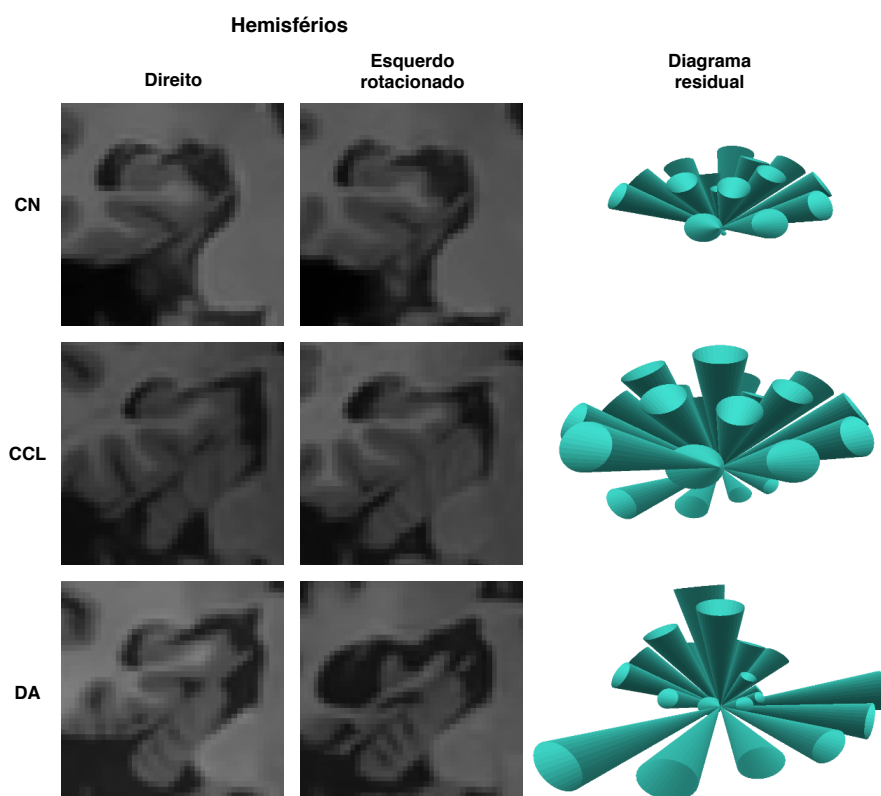


Figura 17 – Ilustração da assimetria hemisférica hipocampal utilizando diagramas de rosas residuais obtidos de pacientes pertencentes aos grupos CN, CCL, DA.

Correlação cruzada normalizada entre os hemisférios

Para medir a semelhança entre os atributos extraídos de cada hemisfério cerebral, a correlação cruzada normalizada entre os vetores de atributos esquerdo (\mathbf{a}_L) e direito (\mathbf{a}_R), ambos de tamanho $N = 26$, de cada um dos momentos estatísticos

foi calculada. Para isso, o vetor direito foi preenchido com zeros em suas duas bordas por $\frac{N}{2}$ unidades, resultando em um novo vetor \mathbf{a}_{R^*} de tamanho $2 \times N$. Em seguida, foi feita a operação de correlação cruzada normalizada entre os vetores \mathbf{a}_L e \mathbf{a}_{R^*} , sendo

$$NCC(x) = \frac{\sum_{i=0}^{N-1} \mathbf{a}_L[i] \cdot \mathbf{a}_{R^*}[i+x]}{\sqrt{\sum_{i=0}^{N-1} \mathbf{a}_L[i]^2} \sqrt{\sum_{i=0}^{N-1} \mathbf{a}_{R^*}[i+x]^2}}, \quad (41)$$

em que $x = 0, 1, \dots, N - 1$.

É importante observar que essa operação ajuda a identificar a natureza da associação entre os dois hemisférios e como eles se correlacionam. Ao usar esta abordagem é possível avaliar o padrão geral de assimetria que cada paciente pode ter, ao invés de comparar isoladamente cada resposta direcional.

4.5.4 Classificação das imagens

Dado que o objetivo desta abordagem é analisar a assimetria entre os hemisférios, os atributos utilizados para classificar as imagens devem conseguir mensurar o quanto os dois hemisférios cerebrais se assemelham ou diferem, o que foi conseguido via a correlação cruzada normalizada calculada para cada medida estatística. A classificação foi realizada utilizando tanto os vetores resultantes da correlação cruzada normalizada das medidas estatísticas individuais, quanto todas as combinações de cada uma das medidas.

4.5.5 Análise estatística dos atributos

Após a classificação das imagens, os atributos assimétricos extraídos foram sumarizados e testados com o teste estatístico ANOVA, com o propósito de verificar se eles possuem diferenças estatisticamente significativas quando comparados aos três grupos de diagnóstico (CN, CCL e DA) e aos gêneros (Masculino e Feminino) e se suas conclusões são condizentes com a literatura médica.

Análise estatística

Para a análise estatística um novo índice de assimetria hipocampal foi proposto (IA),

$$IA = 1 - \frac{1}{N} \sum_{i=0}^{N-1} NCC(x), \quad (42)$$

em que menores valores indicam que os vetores de atributos são bastante similares (caso simétrico) e maiores valores indicam falta de correlação entre os vetores de atributos (caso assimétrico).

Para verificar se existem diferenças estatisticamente significativas entre os valores de IA em cada um dos grupos de diagnóstico (CN, CCL, e DA), o teste estatístico ANOVA foi calculado com $\alpha = 0,05$. Os resíduos dos modelos foram testados para normalidade e homogeneidade da variância. Embora todos os quatro vetores de atributos tenham passado no teste de normalidade, três deles falharam no teste de homogeneidade da variância. Para contornar esse problema, aplicou-se a transformação Yeo-Johnson (YEO; JOHNSON, 2000), sendo

$$\psi(\lambda_b, IA) = \begin{cases} \frac{(IA+1)^{\lambda_b}-1}{\lambda_b}, & \text{if } \lambda_b \neq 0, IA \geq 0 \\ \log(IA + 1), & \text{if } \lambda_b = 0, IA \geq 0 \\ \frac{-[(-IA+1)^{2-\lambda_b}-1]}{(2-\lambda_b)} & \text{if } \lambda_b \neq 2, IA < 0 \\ -\log(-IA + 1) & \text{if } \lambda_b = 2, IA < 0 \end{cases}, \quad (43)$$

em que λ_b foi determinado automaticamente como o valor que maximiza a função de máxima verossimilhança com a biblioteca *scikit-learn*⁵ da linguagem *python*. Após a transformação, todos os quatro vetores de atributos passaram nos testes.

Conforme mencionado na seção 2.11 do Capítulo 2, ao usar o ANOVA para testar a igualdade da média de pelo menos três grupos, não é possível identificar quais grupos diferem. Portanto, para solucionar tal problema, foi aplicado o teste HSD (ABDI; WILLIAMS, 2010) com $\alpha_t = 0,05$.

A seguir é descrita a metodologia utilizada na análise do erro de predição da idade do cérebro. As configurações experimentais, resultados e discussões das análises são apresentados no capítulo seguinte.

⁵ <<https://scikit-learn.org>>

4.6 Análise do erro de predição da idade biológica do cérebro

A Figura 18 ilustra a estrutura utilizada para o desenvolvimento desta abordagem. Primeiramente, as imagens de treinamento foram pré-processadas e as regiões hipocampais, delimitadas pelas máscaras definidas pela etapa de pré-processamento, foram recortadas. A seguir, foram projetados dois experimentos para a tarefa de predição da idade biológica do cérebro. O primeiro utilizou imagens de pacientes com idades entre 20 e 70 anos, e o segundo, pacientes com mais de 70 anos.

Os experimentos foram divididos por faixas etárias para permitir a análise do erro médio de predição da idade do cérebro de pacientes da mesma idade do grupo CN com os pacientes do grupo CCL e DA leve. O primeiro treinamento teve dois objetivos: *i*) servir de base para comparar os resultados obtidos com os modelos existentes e *ii*) fornecer pesos pré-treinados para o segundo, dado que a quantidade de imagens para idades maiores é menor. O segundo treinamento usa os pesos pré-treinados do primeiro modelo, permitindo que seja realizado o experimento com idosos CN. Em seguida, é realizada uma comparação pareada entre as predições de idade dos grupos CN, CCL e DA usando o testes estatísticos ANOVA e Tukey HSD para verificar se existem diferenças estatisticamente significativas entre cada par de grupo de diagnóstico. Ambos os experimentos de treinamento usaram a mesma arquitetura CNN e combinaram as predições de cada hipocampo, esquerdo e direito, para obter a predição final, conforme ilustrado.

4.6.1 Base de imagens

Devido ao uso de aprendizado profundo, esta abordagem utilizou imagens adicionais às mencionadas na seção 4.2. Foram utilizadas 563 imagens da base de dados IXI e 842 da base de dados ADNI, incluindo imagens longitudinais, todas do grupo CN. A Tabela 7 fornece informações sobre a quantidade de imagens usadas em cada experimento.

Como se pode notar, as imagens da base IXI foram utilizadas principalmente no primeiro experimento por conter imagens de pacientes jovens, e as imagens da base ADNI no segundo por conter imagens de pacientes mais idosos. Dado que

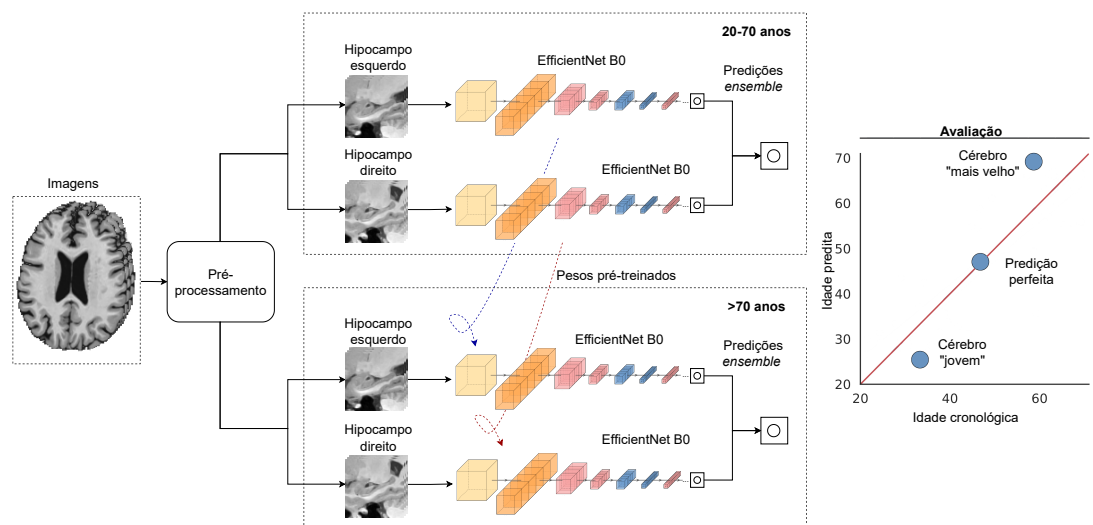


Figura 18 – Fluxograma da sequência de etapas utilizada para o desenvolvimento desta abordagem.

Tabela 7 – Quantidade de imagens (apenas pacientes CN) utilizadas em cada etapa de treinamento.

	Faixa etária	Treinamento		Validação		Teste	
		ADNI/IXI	Total	ADNI/IXI	Total	ADNI/IXI	Total
Imagens de pacientes únicos	20-70	127/449	576	15/29	44	15/28	43
	>70	167/57	224	151/0	151	151/0	151
Inclusão de imagens longitudinais	20-70	238 /449	687	15/29	44	15/28	43
	>70	272 /57	329	151/0	151	151/0	151

nosso objetivo foi explorar as associações de delta de idade do cérebro de pacientes CN de mesma idade com pacientes com DA e CCL, o segundo experimento usou apenas imagens do grupo CN da base ADNI nas divisões de validação e teste. As informações das imagens utilizadas nesta etapa foram apresentadas na Tabela 5, que inclui pacientes CN já contabilizados pela Tabela 7, com a adição de pacientes dos grupos CCL e DA leve.

4.6.2 Definição das regiões de interesse

Nesta etapa foram obtidas as ROIs utilizadas como entrada para a CNN. A partir das imagens com transformação afim, foram recortados *patches* de tamanho

$64 \times 64 \times 64$ ao redor dos hipocampos usando o centro de gravidade das máscaras binárias dos hipocampos, direito e esquerdo, obtidas pelo pré-processamento, como referência.

4.6.3 CNN

Pelo fato da arquitetura das *EfficientNets* terem demonstrado melhores valores de acurácias e eficiência e por sua arquitetura base, B0, possuir um menor número de parâmetros (5,3M) do que outras redes propostas na literatura, ela foi utilizada para a realização desse experimento utilizando os *patches* do hipocampo como entrada.

4.6.3.1 Arquitetura

A arquitetura utilizada nesta pesquisa é composta por 16 blocos, em que a arquitetura da *EfficientNet* B0 foi adaptada para 3D com a alteração de todas as operações 2D para 3D e da última camada para uma função de ativação linear, dessa maneira a saída da rede corresponde à idade biológica predita do cérebro do paciente. O tamanho de entrada corresponde às dimensões dos *patches* do hipocampo, ou seja, $64 \times 64 \times 64 \times 1$, gerando a arquitetura apresentada na Figura 19, com um total de 4 milhões de parâmetros.

Todos os experimentos foram realizados separadamente para os hipocampos esquerdo e direito, e os valores preditos foram combinados (*ensemble*) para a predição final, que corresponde à média das duas predições da rede.

4.6.4 Aumentos dos dados

A base dados ADNI fornece imagens de acompanhamento longitudinal de um mesmo paciente. Essas imagens foram incluídas no treinamento da rede como um “aumento natural” dos dados. Além disso, foram realizados aumentos artificiais que representam a maioria das variações presentes nas imagens de RM. Nesse caso, foram projetadas um total de até vinte transformações, compostas por operações de translação e rotação aplicadas em cada eixo (x , y e z) e com escolhas aleatórias dos parâmetros entre um intervalo definido, ou seja, translação entre ± 10 *pixels*

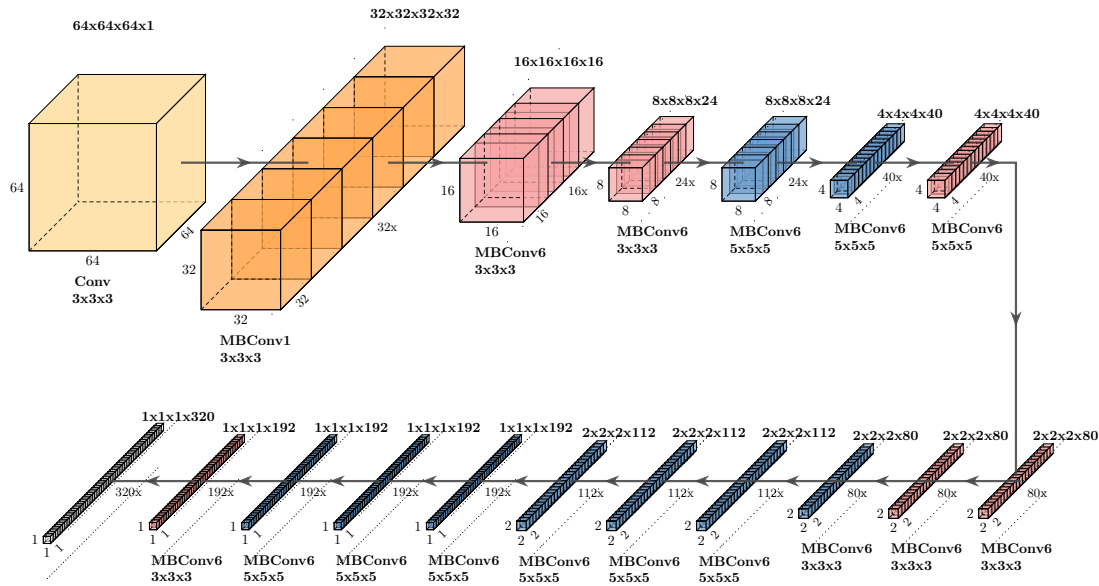


Figura 19 – Arquitetura da *EfficientNet* B0 usada neste trabalho.

e rotação entre ± 20 graus, resultando em seis transformações. Além disso, foi incluído o ruído aditivo Gaussiano com variação aleatória do valor do sigma entre 5 e 30 e o ruído multiplicativo de *bias field* com variação aleatória da magnitude máxima dos coeficientes polinomiais da função de bias entre $-0,3$ e $0,3$, resultando em mais duas transformações. Por fim, as transformações foram combinadas para gerar um total de vinte transformações diferentes, por exemplo, uma rotação aleatória ao longo do eixo x seguida por uma translação aleatória ao longo do eixo z . A Figura 20 ilustra um exemplo de dezesseis imagens aleatórias geradas em um *batch*.

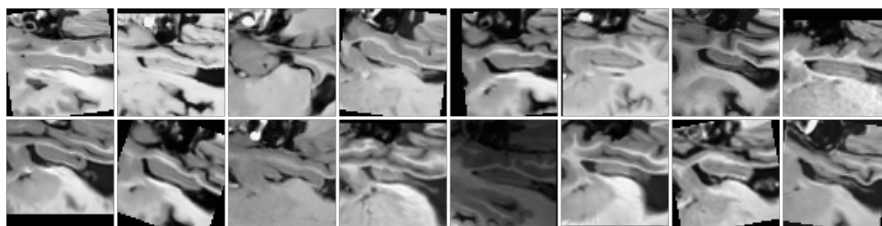


Figura 20 – Exemplo de dezesseis imagens aleatórias de um *batch*.

Após definir as transformações, aumentos proporcionais foram projetados e

incluídos à quantidade de imagens existentes em cada intervalo etário para criar um melhor balanceamento entre as idades. As idades dos pacientes foram estratificadas em intervalos de três anos. As Figuras 21(a) e 21(b) mostram a comparação das distribuições das idades das populações para ambos os experimentos, antes e após o aumento dos dados. Para o primeiro experimento de treinamento, com idades entre 20 e 70 anos, foram obtidas 5.098 imagens de treinamento, e para o segundo, com idades superiores a 70 anos, 1759.

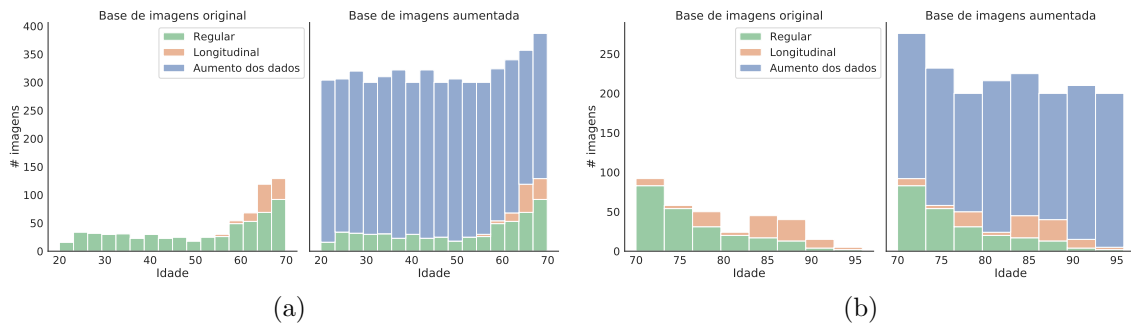


Figura 21 – A distribuição de idade da população de estudo do conjunto de dados originais e do conjunto de dados aumentados. (a) Para o experimento com idade entre 20 e 70 anos, e (b) para o experimento maior que 70 anos.

Todas as transformações foram realizadas com a biblioteca Torchio (PÉREZ-GARCÍA; SPARKS; OURSELIN, 2020)⁶ e aplicadas em tempo real (aumento de dados online) para economizar no armazenamento das imagens e incluir mais variações nos dados, visto que os parâmetros aleatórios geram imagens diferentes em cada época.

4.6.5 Análise estatística

Para a análise estatística foi utilizada a métrica ME (Equação 37), obtida pela diferença entre a idade biológica predita do cérebro e a idade cronológica. Essa métrica é comumente utilizada pela literatura focada na predição da idade do cérebro como um delta (COLE et al., 2017; GASER et al., 2013), portanto o ME será referido ao longo do texto com $\Delta_{\text{Idade cérebro}}$. Sendo assim, os valores

⁶ <https://pypi.org/project/torchio/>

de $\Delta_{\text{Idade cérebro}}$ de cada um dos três grupos de diagnóstico obtidos do segundo experimento foram testados para verificar se existem diferenças estatisticamente significativas. Para isso, foi realizado o teste estatístico ANOVA (LARS; SVANTE, 1989) seguido por uma avaliação pareada entre os grupos com o teste *post-hoc* HSD (ABDI; WILLIAMS, 2010). Os resíduos dos modelos foram testados para normalidade e homogeneidade da variância e passaram nos testes. Além disso, a relação entre os valores do $\Delta_{\text{Idade cérebro}}$ e os valores de MMSE (seção 2.1) foram correlacionados com o teste pareado de Pearson para medir a força e a direção da associação entre às duas variáveis contínuas.

4.6.6 Classificação

Na etapa de classificação o atributo $\Delta_{\text{Idade cérebro}}$ para cada grupo diagnóstico (CN, CCL e DA) foi inserido em um classificador SVM para verificar o seu poder preditivo.

4.7 União das abordagens

Esta etapa consiste em combinar (*ensemble*) as previsões obtidas de cada uma das abordagens desenvolvidas para verificar se elas podem contribuir positivamente para melhorar os resultados individuais obtidos, conforme ilustrado pela Figura 6, no início deste capítulo.

Para isso, foram concatenadas as probabilidades de predição, calculadas com o algoritmo *Calibrated classifier*⁷ da biblioteca sklearn da linguagem *python*, de todas as respostas dos *folds* de teste de cada abordagem. A seguir, a predição foi realizada a partir da média ponderada das probabilidades das previsões de cada abordagem, calculada conforme a fórmula:

$$\text{Prob}_{\text{final}} = \frac{(\text{Prob} * \text{Peso})_{\text{atrofia}} + (\text{Prob} * \text{Peso})_{\text{assimetria}} + (\text{Prob} * \text{Peso})_{\text{delta}}}{\text{Peso}_{\text{atrofia}} + \text{Peso}_{\text{assimetria}} + \text{Peso}_{\text{delta}}}, \quad (44)$$

nesse caso, cada probabilidade foi ponderada por um diferente peso, que define a contribuição de cada abordagem. Os pesos foram definidos com uma busca de

⁷ <https://scikit-learn.org/stable/modules/generated/sklearn.calibration.CalibratedClassifierCV.html>

parâmetros usando a AUC como critério. A AUC e o ponto de corte da curva ROC foram obtidos usando a probabilidade final ($\text{Prob}_{\text{final}}$). O ponto de corte foi selecionado como o ponto sobre a curva ROC mais próximo do canto superior esquerdo do gráfico, ponto $(0, 1)$. Esse ponto de corte, usado também para o cálculo das demais métricas usadas na avaliação dos resultados, garante uma alta sensibilidade e especificidade para o sistema.

4.8 Comparação dos resultados

Para comparar os resultados obtidos pelas abordagens propostas, um método baseado em regiões de interesse (*ROI-based*) foi implementado. O método *ROI-based* é comumente usado como o “método controle” em muitos artigos publicados na literatura (ZHANG et al., 2016; ZHANG et al., 2017; LIU et al., 2020; LIU et al., 2017; LIU et al., 2018; LIAN et al., 2020).

Para a sua implementação foi utilizada a imagem do cérebro de um único paciente, fornecida pelo *Montreal Neurological Institute* (MNI) (COLLINS et al., 1998), em conjunto com o AAL proposto por Tzourio-Mazoyer et al. (2002), que fornece a segmentação de noventa ROIs predefinidas do cérebro para esse paciente da MNI. Primeiro, transformações deformáveis, que mapeiam as móveis (imagens da base ADNI) para uma imagem fixa (imagem da base MNI), foram obtidas usando o algoritmo de registro deformável do NiftyReg. Em seguida, cada transformação foi aplicada às noventa ROIs do AAL para seu posicionamento na respectiva imagem de estudo. Depois disso, as imagens da ADNI foram segmentadas em três tipos diferentes de tecidos (GM, WM e CSF) utilizando a técnica FAST (ZHANG; BRADY; SMITH, 2001). Então, os volumes dos tecidos GM, WM e CSF foram calculados dentro de cada uma das noventa ROIs, em todas as imagens de estudo. Finalmente, os volumes GM foram normalizados pelo volume intracraniano total e utilizados como atributos para a classificação. Nesse caso, o volume intracraniano total foi calculado como a soma dos volumes GM, WM e CSF.

4.9 Considerações finais

Este capítulo apresentou informações sobre as bases de dados utilizadas, os algoritmos de pré-processamento e as metodologias utilizadas para o desenvolvimento de cada uma das abordagens. Além disso, foi apresentada a metodologia utilizada para a implementação de um “método controle”, desenvolvido para a comparação dos resultados. Os resultados e configurações experimentais obtidos e utilizados para cada abordagem são apresentados e discutidos no Capítulo 5.

Capítulo 5

Resultados e discussões

Este capítulo apresenta as configurações experimentais, os resultados e discussões obtidos para as três abordagens propostas nesta pesquisa.

5.1 Análise de atrofia estruturais

A metodologia utilizada nessa abordagem foi descrita na seção 4.4 e as informações sobre as imagens dos pacientes utilizados foram apresentadas na Tabela 5 da subseção 4.2.2. As configurações experimentais e os resultados obtidos para a classificação das imagens são apresentados nas subseções que seguem.

5.1.1 Configurações experimentais

A metodologia proposta para a classificação das imagens de RM nos casos CN×CCL, CCL×DA, e CN×DA para os dois níveis foi validada usando duas divisões de validação cruzada estratificada. Esse procedimento teve como objetivo diminuir as chances de *overfitting* e garantir que as imagens utilizadas na busca dos parâmetros não fossem utilizadas para a avaliação do método. Para isso, foi utilizada uma abordagem semelhante à validação cruzada aninhada, com a diferença de que foram efetuadas duas divisões estratificadas separadas, uma com cinco e

outra com dois *folders*. A Figura 22 mostra a divisão dos *folders*, com o número de imagens utilizadas para cada nível e as ilustrações das divisões.

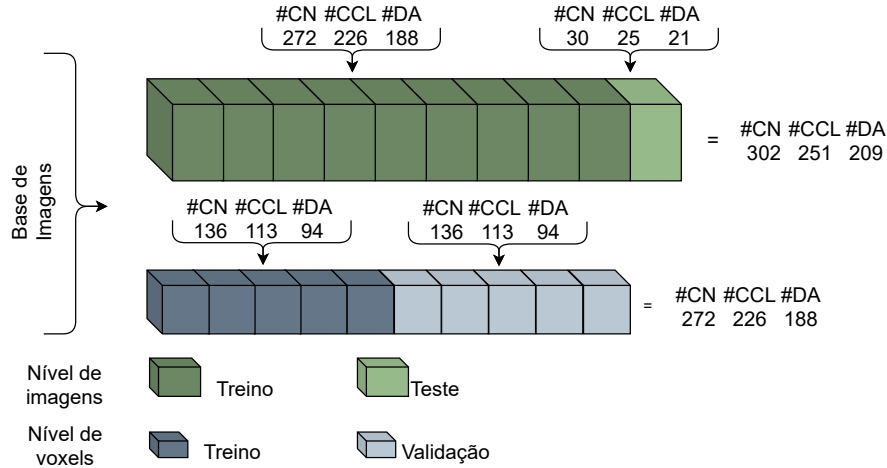


Figura 22 – Divisão dos *folders* utilizada para a metodologia de dois níveis de classificação.

Foi utilizada uma divisão em dez *folders* (cores verdes) para o nível da imagem e uma divisão em dois *folders* (cores azuis) para o nível de pontos salientes. As imagens de treinamento (verde-escuro), pertencentes aos dez *folders*, foram divididas em dois *folders* e utilizadas, respectivamente, para treinamento (azul-escuro) e validação (azul-claro) no nível de pontos salientes. As imagens de teste (verde-claro), no final dos dez *folders*, foram usadas apenas para a predição usando os modelos aprendidos.

Para avaliação do desempenho, foram utilizadas cinco métricas, a AUC, acurácia, sensibilidade, especificidade e F1-score. Nos experimentos, as imagens de RM de pacientes com doenças cognitivas foram consideradas como amostras positivas e as imagens de pacientes CN como negativas. Para o experimento comparando casos de DA e CCL, os casos de CCL foram considerados como positivos e os com DA como negativos. Além disso, devido ao conjunto de dados ser levemente desbalanceado, os pesos do modelo foram automaticamente ajustados para serem inversamente proporcionais às frequências das classes dos dados de entrada usando o parâmetro *class_weight*¹ da biblioteca *scikit-learn*² da linguagem *python*.

¹ <http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

² <https://scikit-learn.org>

Nível de pontos salientes

Para a classificação dos descritores dos pontos salientes foi realizado um *grid-search* utilizando validação cruzada estratificada com 5 *folds* para determinar os hiperparâmetros dos SVMs. Em relação ao *kernel*, foram testados o RBF, o polinomial e o linear. Os *kernels* RBF e polinomial possuem o parâmetro γ que ajusta o conjunto de dados de treinamento; se esse valor for muito grande, ele pode causar *overfitting*, e se for muito pequeno, o modelo será muito limitado para capturar a complexidade dos dados (NELL; SHAWE-TAYLOR, 2000). Além disso, o *kernel* polinomial também possui os parâmetros k e d , que correspondem respectivamente à homogeneidade do *kernel* e o grau do polinômio. Se $k = 0$, então o *kernel* é homogêneo (todos os seus monômios possuem o mesmo grau) e se $k > 0$, então o *kernel* é não homogêneo (todos os monômios diferem e possuem até d graus). Se os parâmetros do *kernel* polinomial forem ajustados para $d = 1$, $\gamma = 1$, e $k = 0$, então, o *kernel* polinomial pode ser utilizado como um SVM linear. Neste trabalho foi utilizado o *kernel* polinomial homogêneo ($k = 0$), o valor do grau foi variado entre 1 e 3 e o valor de γ foi definido automaticamente pela fórmula

$$\gamma = \frac{1}{M_x \cdot \sigma_x}, \quad (45)$$

em que M_x e σ_x correspondem, respectivamente, à dimensionalidade e à variância das amostras de treinamento.

Os valores de C para os três *kernels* do SVM foram determinados experimentalmente utilizando *coarse-to-fine-grid search*. Nesse caso, o parâmetro C foi variado entre $[2^{-5}; 2^5]$ com passos definidos por uma base 2 com o expoente incrementado em 0,5, ou seja, $2^{-5}, 2^{-4.5}, \dots, 2^5$; seguido por uma pesquisa mais fina (*fine-search*) na vizinhança do melhor parâmetro obtido pelo *coarse-grid*, \tilde{C} , com um intervalo de valores entre $[(\log_2 \tilde{C}) - 2, (\log_2 \tilde{C}) + 2]$ e passos definidas por uma base 2 com o expoente incrementado em 0,25.

Neste nível, conforme mencionado na subseção 4.4.3, cada descritor de ponto foi considerado como uma amostra individual, portanto, não foi considerada qualquer relação que possa existir com os outros descritores extraídos da mesma imagem. No entanto, após alguns experimentos iniciais, notou-se que o classificador estava aprendendo padrões específicos dos pontos e, conseqüentemente, perdendo

informações globais da imagem. Portanto, a informação da imagem de origem dos descritores foi utilizada apenas para selecionar os hiperparâmetros do classificador. Nesse caso, foi calculado um valor de F1 médio por imagem a partir do valor de F1 individual de cada descritor. Essa métrica foi implementada usando a função *group k-fold*³ fornecida na biblioteca *scikit-learn* — cada imagem foi considerada um grupo e a classe da imagem em análise foi sempre considerada positiva. Assim, dada a saída do classificador, todas as amostras de descritores com a mesma classe da imagem foram consideradas positivas e as outras negativas.

Embora esse procedimento de pontuação tenha custado algumas classificações erradas de pontos salientes, foram obtidos parâmetros que maximizaram o padrão de imagem ao invés dos descritores individualmente. A razão para isso é que se o conjunto de parâmetros que maximiza apenas os descritores tivesse sido selecionado, poderia ter sido treinado um classificador que eventualmente acertaria ou erraria todos os pontos salientes de algumas imagens, o que afetaria negativamente o próximo nível de classificação.

Nível de imagens

Para a classificação das imagens foi realizado um *grid-search* utilizando validação cruzada estratificada com 10 *folds* para determinar os hiperparâmetros dos SVMs (ou seja, *kernels* lineares, RBF e polinomiais), considerando a AUC como um critério de desempenho e γ como $\frac{1}{M_x \sigma_x}$. O parâmetro C que maximiza a AUC de cada classificador SVM foi determinado em um *coarse-grid* com um intervalo de valores entre $[2^{-5}; 2^{10}]$ e passos definidos por uma base 2 com o expoente incrementado por 0, 25, ou seja, $2^{-5}, 2^{-4.75}, \dots, 2^{10}$; seguido por uma pesquisa mais fina (*fine-search*) na vizinhança do melhor parâmetro obtido pelo *coarse-grid*, \tilde{C} , com um intervalo de valores entre $[(\log_2 \tilde{C}) - 2, (\log_2 \tilde{C}) + 2]$ e passos definidos por uma base 2 com o expoente incrementado em 0, 125.

5.1.2 Análise dos parâmetros

Os parâmetros utilizados em cada uma das etapas propostas nesta abordagem foram avaliados a seguir.

³ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GroupKFold.html

5.1.2.1 Detecção de pontos salientes 3D

Conforme mencionado na seção 2.7, os pontos salientes foram detectados utilizando a técnica de congruência de fase 3D implementada via convolução de uma imagem com os filtros log-Gabor 3D. Os valores dos parâmetros do banco de filtros foram ajustados de maneira que a técnica de congruência de fase realçasse as bordas contidas nas regiões de interesse e, assim, permitisse obter pontos mais estáveis, ou seja, menos suscetíveis ao ruído. Portanto, os parâmetros do banco de filtros 3D log-Gabor foram definidos como: níveis de escala $N_s = 4$, ângulos azimutes $N_{a,0} = 6$, ângulos de elevação $N_e = 3$, frequência máxima central $\omega_{\max} = 0,33$, fator de escala $\gamma = 2,1$, largura de banda do filtro $\eta_\beta = 0,55$, e desvio angular $\sigma_\alpha = 1,2$.

Para evitar respostas acentuadas dos filtros nas bordas das imagens filtradas, a congruência de fase foi aplicada a toda a região do encéfalo, mas, a região de detecção dos pontos salientes foi restrita pelas máscaras binárias (ROI) das estruturas hipocampais. Cabe ressaltar que a restrição da região imposta à detecção dos pontos salientes não foi aplicada no momento do cálculo dos descritores dos pontos salientes, permitindo, assim, um alcance maior para melhor caracterização dos pontos. Nesse caso, um raio máximo (r_{max}) foi definido para a esfera representando a região do descritor. Além disso, foi utilizado um limite de saliência de 0,01 (limiar T_λ da Equação 15), que define se um determinado ponto será considerado como detectado, para atingir uma quantidade razoável de pontos nas regiões dos hipocampus.

A Figura 23 ilustra a vista coronal do cérebro como uma representação visual dos dois autovalores construídos a partir da congruência de fase, utilizados para a detecção e descrição dos pontos salientes. Como referência para os mapas, na Figura 23(a) é apresentada a vista coronal em T1- w , na Figura 23(b) é ilustrado o maior autovalor, λ_0 , utilizado como mapa de bordas para a descrição dos pontos salientes e na Figura 23(c) o menor autovalor, λ_2 , utilizado como mapa de saliência para a detecção de pontos salientes.

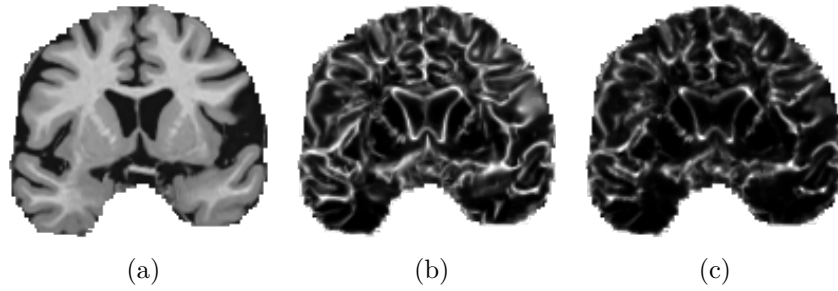


Figura 23 – Representação visual (vista coronal) da imagem em (a) T1- w e do (b) maior e (c) menor autovalor do cérebro resultantes da congruência fase.

5.1.2.2 Descritores

O melhor conjunto de parâmetros para os descritores foram selecionados nos experimentos a seguir.

Casamento

Os descritores usados para o casamento de pontos salientes foram avaliados em cinquenta imagens de RM 3D moderadamente deformadas. A Figura 24 mostra o diagrama de etapas para este experimento.

Primeiro, os pontos salientes foram detectados nas regiões hipocampais esquerda e direita da imagem de referência da NAC (P_{NAC}). Em seguida, transformações deformáveis aleatórias foram aplicadas na mesma imagem, gerando cinquenta imagens moderadamente deformadas ($Im_1, Im_2, \dots, Im_{50}$). Então, os pontos salientes com seus respectivos descritores ($P_{Im_1}, P_{Im_2}, \dots, P_{Im_{50}}$) foram detectados em cada imagem deformada, e a transformação correspondente T_i foi aplicada às coordenadas dos pontos detectados para definir suas localizações equivalentes nas imagens deformadas. Assim, foi gerado um conjunto de correspondências esperadas (*ground-truths*: $GT_{P_{Im_1}}, GT_{P_{Im_2}}, \dots, GT_{P_{Im_{50}}}$) para todos os pontos das imagens deformadas em relação aos da imagem de referência. Em seguida, os pontos salientes das cinquenta imagens deformadas foram casados ($C_{P_{Im_1}}, C_{P_{Im_2}}, \dots, C_{P_{Im_{50}}}$) com os da imagem da NAC. Finalmente, a distância eu-

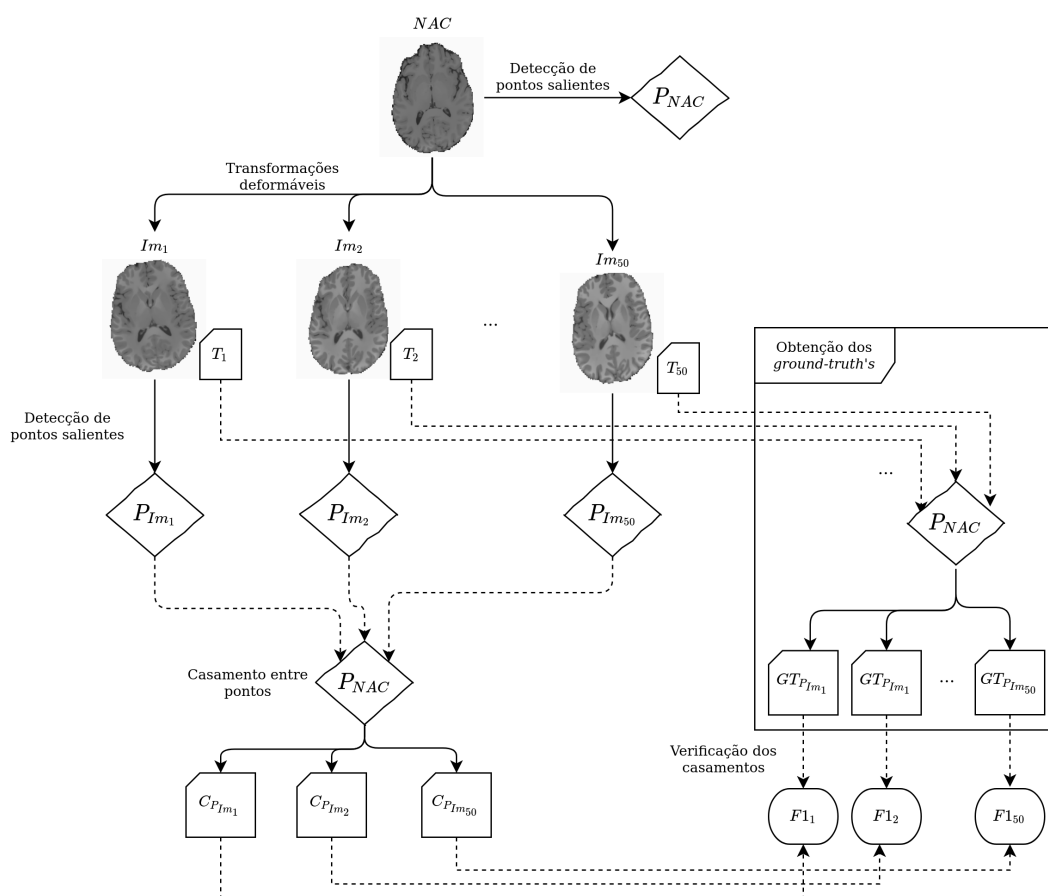


Figura 24 – Sequência de etapas utilizadas para a avaliação dos parâmetros dos descritores.

clidiana entre os pontos casados e os *ground-thruths* foi verificada e a métrica F1-score foi calculada.

O melhor conjunto de parâmetros dos descritores foi escolhido considerando o maior valor médio de F1-score obtido para as regiões dos hipocampos. Foi utilizado um erro de tolerância de 7 voxels de distância euclidiana, visto que a detecção de um ponto equivalente pode não ocorrer na localização estimada predeterminedada exata. O parâmetro do descritor η_r , (correspondente à distância radial) foi testado com os valores de 3, 4 e 5, os ângulos de azimute relativos, η_a , com valores de 8, 10 e 12; os ângulos de elevação relativos, η_e , sempre corresponderam à metade dos valores de η_a ; e o raio máximo, r_{max} , foi variado com valores de 16, 32 e 64 voxels.

Os resultados dos experimentos são mostrados na Figura 25. Nesse caso, cada F1-score corresponde ao valor médio das regiões hipocampais direita e esquerda obtidas para os descritores 3D global (pontilhado) e local (contínuo). Como pode ser visto, o melhor desempenho do descritor foi alcançado para $r_{max} = 32$ (cor laranja), seguido por $r_{max} = 64$ (verde) e $r_{max} = 16$ (azul). Além disso, o descritor local teve um desempenho melhor do que o global para todos os valores de r_{max} testados. O melhor resultado, pontuação F1 de 0,885, foi obtido para $r_{max} = 32$ usando o descritor local com os parâmetros $\eta_a = 10$, $\eta_e = 05$ e $\eta_r = 05$. Portanto, este conjunto de valores de parâmetro foi usado para o casamento de pontos salientes.

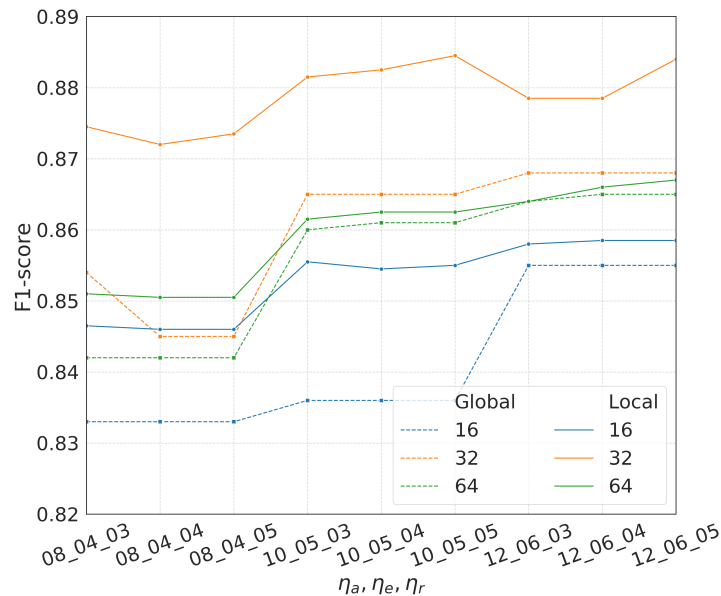


Figura 25 – Influência dos parâmetros do descritor no casamento entre os pontos salientes.

Classificação

Os descritores utilizados para a classificação de pontos salientes foram avaliados na região hipocampal direita de quarenta imagens de RM (vinte CN e vinte DA) não incluídas no grupo de teste e usando a concatenação dos atributos dos pontos (ou seja, GM|WM|CSF).

O melhor conjunto de parâmetros foi definido considerando o maior valor de AUC obtido da classificação e os seguintes valores de parâmetros foram testados: as distâncias radiais, $\eta_r = 3, 4, \text{ e } 5$; os ângulos de azimute relativos, $\eta_a = 8 \text{ e } 10$; os ângulos de elevação relativos, $\eta_e = \text{ metade dos valores de } \eta_a$; e o raio máximo, $r_{max} = 16 \text{ e } 32$.

Os resultados são apresentados na Figura 26, sendo que os resultados para os descritores locais e globais são representados por linhas contínuas e pontilhadas, respectivamente. Como pode ser observado, para este experimento, o descritor global teve um desempenho melhor do que o descritor local para quase todos os conjuntos de parâmetros. Além disso, o descritor global teve melhor desempenho para $r_{max} = 32$ do que $r_{max} = 16$. O melhor resultado, com pontuação F1 de 0,795, foi obtido para $r_{max} = 32$ usando o descritor global com os parâmetros $\eta_a = 08, \eta_e = 04$ e $\eta_r = 03$. Portanto, esse conjunto de parâmetros foi usado para a classificação no nível de pontos salientes.

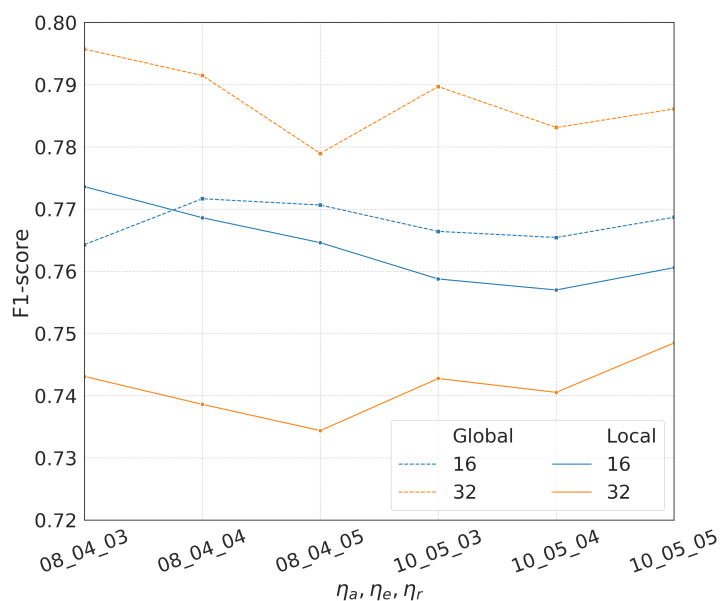


Figura 26 – Influência dos parâmetros do descritor na classificação dos pontos salientes.

5.1.2.3 Seleção dos pontos salientes médios do atlas

Foram construídos dois atlas de pontos salientes usando as imagens do banco de dados IXI, um para a região hipocampal direita com 43 pontos e outro para a esquerda com 47. Para a detecção e descrição de seus pontos salientes, foram utilizados os mesmos valores de parâmetro mencionados na subseção 5.1.2.1. Para a construção do atlas, foram utilizados os mesmos parâmetros sugeridos pelos autores e definidos por padrão na biblioteca ITK (subseção 2.8). Além disso, foi utilizada uma região vizinha de raio de 5 *voxels* e uma densidade mínima de 0,25 para encontrar a posição máxima.

A Figura 27 ilustra um exemplo de resultado obtido para o mapa de densidade do atlas probabilístico construído para a região do hipocampo esquerdo, com o mapa de densidade sobreposto à imagem da NAC. A intensidade de cada voxel no mapa de densidade é proporcional a maior (amarelo) ou menor (vermelho) presença de pontos salientes naquela região espacial. Nota-se que, embora possua baixo contraste, existem fortes picos nas proximidades das bordas hipocampais.

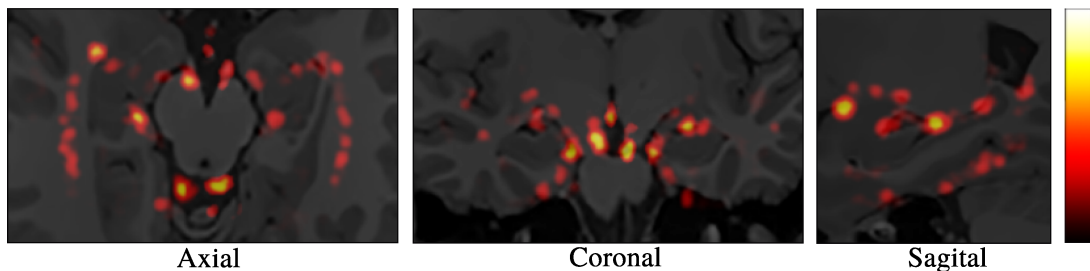


Figura 27 – Atlas probabilístico de pontos salientes 3D da região hipocampal esquerda.

Como os atlas são construídos a partir de cérebros saudáveis, e os experimentos CN×CCL e CN×DA contêm cérebros saudáveis, foi utilizado um nível de significância de 0,01 como limite para a seleção dos pontos do atlas. No entanto, visto que o experimento CCL×DA não contém uma classe saudável e o cérebro CCL é mais semelhante aos atlas saudáveis do que o com DA, foi utilizado um nível de significância de 0,05 como limite para a seleção dos pontos do atlas. Um nível de significância mais restrito foi utilizado para os dois primeiros experimentos, CN×CCL e CN×DA, porque as imagens de indivíduos CN são mais semelhantes

às imagens do atlas de pacientes saudáveis do que as dos grupos comparados, permitindo, assim, que seja obtido um número razoável de pontos após a seleção. No entanto, no último experimento (CCL×DA), as imagens do grupo CCL são mais semelhante as do atlas de pacientes saudáveis do que as do grupo DA, mas não é tão semelhante as dos próprios grupos CN, portanto, se o mesmo nível de significância fosse utilizado, resultaria em um pequeno número de pontos selecionados, impossibilitando a realização deste último experimento.

Como os experimentos foram realizados com validação cruzada, foram produzidos mapas de *p-valor* distintos para cada experimento e *fold*, o que resultou em números diferentes de pontos salientes selecionados para cada um. Para o experimento CN×CCL foram selecionados 27, 8 ± 0 , 87 e 29, 7 ± 1 , 9 pontos salientes para as regiões hipocampal direita e esquerda, respectivamente. Para o experimento CCL×DA foram selecionados 27, 4 ± 2 , 42 e 29, 6 ± 2 , 37 pontos, e para o CN×DA, os números médios, com os respectivos desvios-padrão, dos pontos selecionados foram 38, 9 ± 1 , 14 e 44, 9 ± 0 , 94. Pelo fato dos pontos terem sido selecionados com o auxílio do atlas de pontos salientes, construído a partir de imagens de cérebro saudáveis, é possível notar que o número de pontos selecionados foi maior para os experimentos contendo o grupo CN (ou seja, CN×DA e CN×CCL). Além disso, foram selecionados mais pontos para o CN×DA do que para o CN×CCL, visto que, devido às fases do continuum da DA, mencionadas no Capítulo 2, as diferenças entre cérebros dos grupos CN e DA apresentam mudanças mais distintas em comparação com os cérebros dos grupos CN e CCL.

5.1.3 Resultados

Os resultados para os dois níveis de classificação em conjunto com avaliações qualitativas são apresentados a seguir.

5.1.3.1 Atributos dos pontos salientes

Os atributos dos pontos salientes foram extraídos conforme a metodologia descrita na subseção 4.4.3. Para o primeiro experimento de classificação, foi utilizado o histograma log-esférico global com os seguintes valores de parâmetro: $\eta_a = 8$, $\eta_e = 4$, $\eta_r = 3$, e $r_{max} = 32$. Uma análise detalhada da seleção desses parâmetros

foi realizada na seção 5.1.2. Ao definir esses valores de parâmetros para o descritor, o tamanho dos descritores de pontos salientes usados como vetores de atributos para classificação foi de noventa e seis, exceto o caso WM|GM|CSF, que foi de três vezes esse tamanho.

Além disso, os melhores hiperparâmetros do classificador de pontos foram encontrados usando os descritores do conjunto de imagens de treinamento, com validação cruzada em cinco *folds*, para maximizar o F1-score médio por imagem (explicado na subseção 5.1.1). Em seguida, os modelos foram treinados e a predição realizada para os conjuntos de validação e teste. Como o conjunto de dados foi dividido usando validação cruzada de dez *folds* no nível da imagem, esse experimento foi executado dez vezes. Embora diferentes *kernels* do SVM tenham sido testados, os melhores resultados foram obtidos usando o *kernel* RBF e, portanto, os resultados da classificação são apresentados e discutidos apenas para este *kernel*. Os resultados médios de classificação e seus respectivos desvios-padrão, para todos os experimentos (ou seja, CN×CCL, CCL×DA e CN×DA) correspondentes à maior AUC obtida para as imagens de validação com dez *folds* no nível de pontos salientes, são mostrados na Tabela 8. Os melhores resultados (para cada experimento) estão indicados em negrito.

Tabela 8 – Resultados da AUC média±desvio-padrão para os experimentos CN×CCL, CCL×DA, e CN×DA realizados no nível de pontos salientes.

	CN×CCL		CCL×DA		CN×DA	
	hip. direito	hip. esquerdo	hip. direito	hip. esquerdo	hip. direito	hip. esquerdo
WM	0,66 ± 0,01	0,70 ± 0,01	0,62 ± 0,01	0,60 ± 0,01	0,79 ± 0,004	0,77 ± 0,007
GM	0,71 ± 0,01	0,66 ± 0,01	0,61 ± 0,01	0,62 ± 0,02	0,82 ± 0,005	0,80 ± 0,004
CSF	0,75 ± 0,02	0,74 ± 0,01	0,58 ± 0,01	0,63 ± 0,01	0,86 ± 0,01	0,86 ± 0,004
WM GM CSF	0,77 ± 0,02	0,74 ± 0,01	0,62 ± 0,01	0,65 ± 0,01	0,88 ± 0,005	0,86 ± 0,01

A partir dos resultados apresentados na Tabela 8 é possível verificar que a proposta de concatenação de atributos influenciou positivamente nos resultados dos três experimentos. Além disso, os atributos CSF resultaram nos segundos melhores resultados, seguidos pelo GM e pelo WM. Os melhores valores de AUC nos experimentos CN×CCL e CN×DA foram obtidos principalmente para a região direita do hipocampo em quase todos os atributos. O mesmo comportamento

aconteceu para a região do hipocampo esquerdo no experimento CCL×DA. Esta metodologia de dois níveis visou aumentar o número de amostras de treinamento no primeiro nível e criar uma “margem de segurança” que permite a ocorrência de erros de classificação no nível de pontos salientes, que podem ser desconsiderados no nível da imagem.

5.1.3.2 Classificação das imagens

Para o segundo experimento, os melhores hiperparâmetros do classificador foram determinados via maximização dos valores de AUC obtidos do processamento do conjunto de imagens de validação, com validação cruzada em dez *folders*. Em seguida, os modelos foram treinados usando as imagens de validação e as predições foram realizadas usando as imagens de teste. Embora diferentes *kernels* SVM tenham sido testados, os melhores resultados foram obtidos usando o *kernel* polinomial com grau um, portanto, os resultados de classificação são apresentados e discutidos apenas para este *kernel*. Além disso, conforme o experimento anterior, esse experimento também foi executado dez vezes. Os resultados da classificação para os experimentos CN×CCL, CCL×DA, e CN×DA são mostrados nas Tabelas 9, 10, e 11, respectivamente. Eles correspondem à média e desvio padrão da AUC, acurácia, sensibilidade, especificidade e F1-score obtidos para as imagens de teste de todos os *folders*. Os melhores resultados para cada experimento estão em negrito.

Tabela 9 – Resultados de classificação dos 10 *folders* de validação cruzada (média±desvio-padrão) para o experimento CN×CCL.

	Atributos			
	WM	GM	CSF	WM GM CSF
AUC	0,77 ± 0,04	0,80 ± 0,06	0,82 ± 0,04	0,83 ± 0,05
Acc(%)	70,35 ± 4,52	74,7 ± 5,3	74,87 ± 3,44	75,58 ± 3,6
Sens(%)	69,69 ± 9,23	70,57 ± 9,76	74,14 ± 7,03	72,94 ± 8,41
Espec(%)	70,87 ± 7,66	78,12 ± 8,36	75,51 ± 6,13	77,81 ± 6,84
F1	0,68 ± 0,06	0,72 ± 0,06	0,73 ± 0,04	0,73 ± 0,04

Para os experimentos CN×CCL, foi obtido um valor de AUC de 0,83 e uma acurácia de 75,58%. Além disso, os valores de sensibilidade e especificidade foram muito semelhantes ($\approx 5\%$ de diferença) e um desvio padrão relativamente pequeno

(mínimo de 4% e máximo de 8,41%) foi obtido entre os dez *folders* para todas as métricas avaliadas.

Tabela 10 – Resultados de classificação dos 10 *folders* de validação cruzada (média±desvio-padrão) para o experimento CCL×DA.

	Atributos			
	WM	GM	CSF	WM GM CSF
AUC	0,7 ± 0,09	0,66 ± 0,15	0,72 ± 0,07	0,73 ± 0,07
Acc(%)	67,64 ± 8,08	66,75 ± 6,03	66,99 ± 8,34	69,8 ± 6,73
Sens(%)	74,2 ± 11,75	73,74 ± 12,95	66,15 ± 10,45	74,09 ± 6,3
Espec(%)	59,77 ± 10,12	58,36 ± 17,47	67,91 ± 8,69	64,57 ± 11,17
F1	0,71 ± 0,08	0,71 ± 0,06	0,68 ± 0,09	0,73 ± 0,06

Para o experimento CCL×DA, foi obtida uma AUC de 0,73 e uma acurácia de 69,8%. Aqui, os valores de sensibilidade e especificidade ficaram um pouco mais distantes ($\approx 10\%$ de diferença). Semelhante ao caso CN×CCL, o desvio padrão entre os dez *folders* para todas as métricas avaliadas permaneceu pequeno (mínimo de 6% e máximo de 11,17%).

Tabela 11 – Resultados de classificação dos 10 *folders* de validação cruzada (média±desvio-padrão) para o experimento CN×DA.

	Atributos			
	WM	GM	CSF	WM GM CSF
AUC	0,9 ± 0,03	0,94 ± 0,03	0,94 ± 0,04	0,95 ± 0,03
Acc(%)	80,82 ± 4,19	87,47 ± 3,96	86,3 ± 4,12	89,24 ± 4,04
Sens(%)	80,86 ± 3,7	85,1 ± 6,52	83,67 ± 7,97	85,58 ± 8,5
Espec(%)	80,79 ± 8,54	89,04 ± 6,0	88,08 ± 6,0	91,72 ± 5,0
F1	0,78 ± 0,03	0,85 ± 0,05	0,83 ± 0,05	0,87 ± 0,05

Finalmente, para o experimento CN×DA foi obtido um valor de AUC de 0,95 e uma acurácia de 89,24%. Os valores gerais de desvio padrão de todas as métricas avaliadas também foram pequenos (mínimo de 3% e máximo de 8,5%) entre os dez *folders*.

A Figura 28 mostra as curvas ROC com os valores AUC correspondentes para a classificação CN×CCL, CCL×DA, e CN×DA. A partir das curvas, observa-se a relação entre a sensibilidade e a especificidade em diferentes limiares. Com esses

valores, os especialistas do domínio podem estudar e selecionar diferentes limites para as métricas para determinar o ponto operacional do sistema.

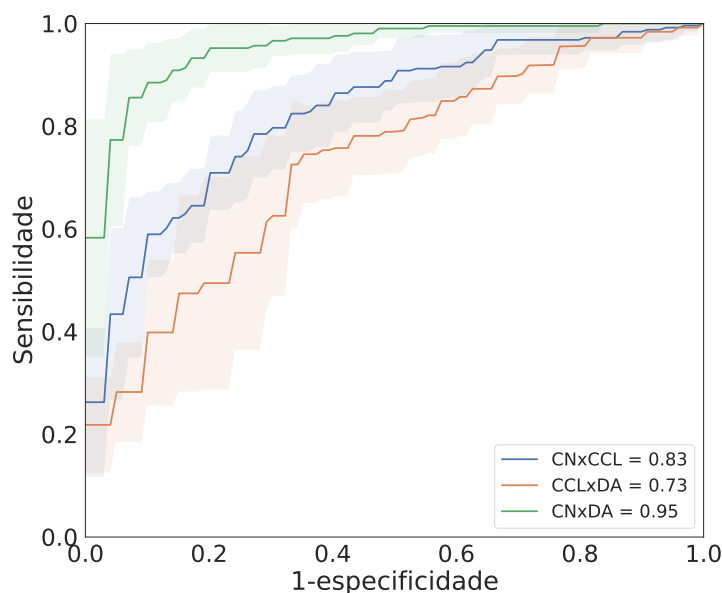


Figura 28 – Curvas ROC e valores de AUC correspondentes aos resultados de classificação de CN×CCL, CCL×DA e CN×DA.

Os resultados apresentados nas Tabelas 9, 10 e 11 mostraram a mesma tendência geral obtida dos resultados do nível de pontos salientes, ou seja, a concatenação de atributos influenciou positivamente os resultados, seguida pelos atributos CSF, GM e WM. Esta tendência foi consistente em todas as etapas desta abordagem, visto que os experimentos do nível de pontos foram realizados separadamente para cada região hipocampal e seus descritores classificados sem usar nenhuma relação com aqueles extraídos da mesma imagem e, em seguida, todas as informações foram concatenadas para o nível da imagem. Além disso, devido à morte de células neuronais causada pela DA, o GM, que é o principal tecido das regiões do hipocampo, encolhe e é substituído pelo CSF (GUADALUPE et al., 2014; WOOLARD; HECKERS, 2012). Portanto, os resultados obtidos são consistentes com as etapas da DA, visto que os atributos CSF e GM apresentaram resultados de classificação superiores aos do WM.

Comparando os resultados dos dois níveis de classificação, nota-se consistência entre as conclusões e um aumento dos valores de AUC no nível da imagem. Sendo

assim, conforme a hipótese inicialmente formulada, a estrutura em dois níveis criou uma “margem de segurança” para a classificação das imagens de RM.

Análise visual qualitativa

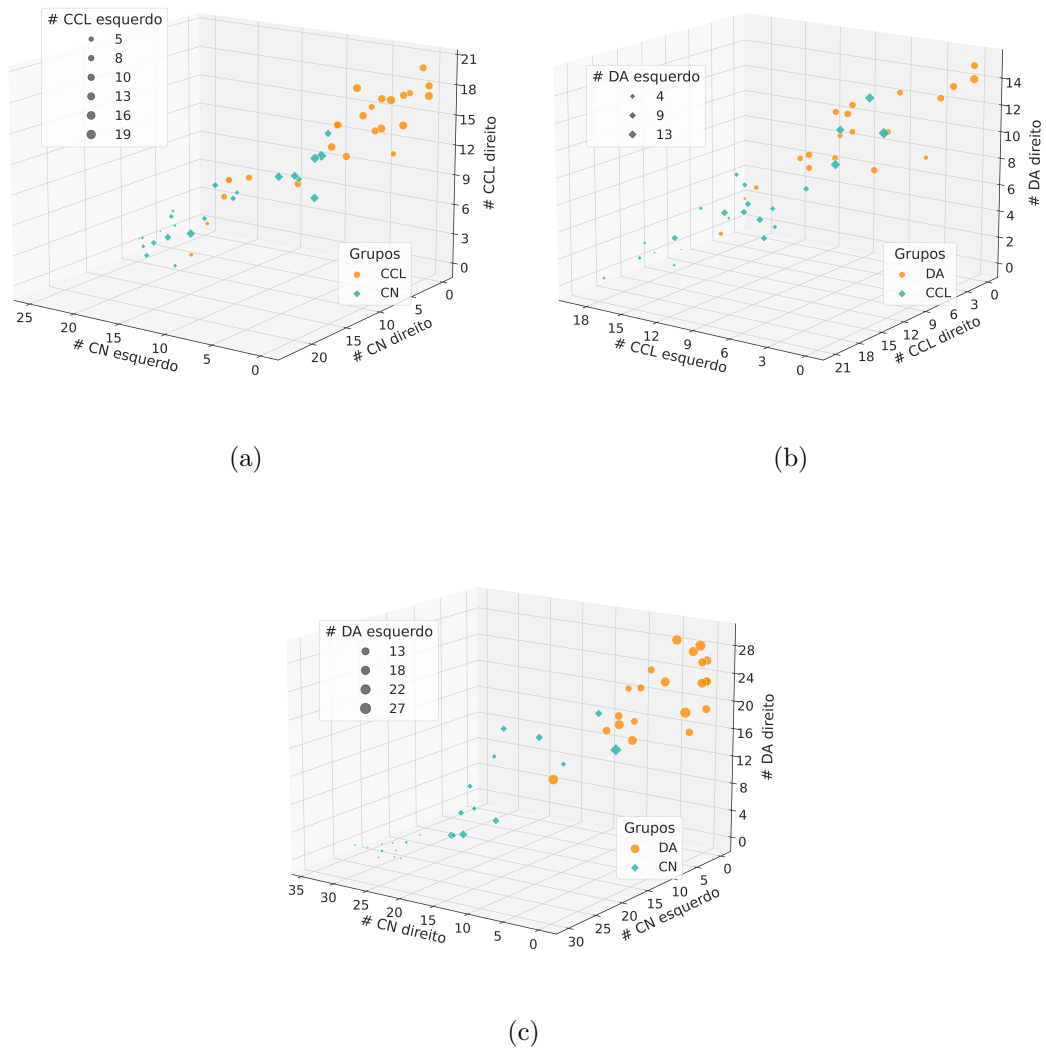


Figura 29 – Gráficos de dispersão de quatro atributos (de um total de dez) correspondentes a um *fold* de cada um dos experimentos: CN×CCL, CCL×DA, e CN×DA, respectivamente.

As Figuras 29 (a), (b) e (c) mostram os gráficos de dispersão de quatro (de dez) atributos de um *fold* dos experimentos CN×CCL, CCL×DA, e CN×DA, respectivamente, para ilustrar a interpretabilidade dos vetores de atributos construídos para o nível da imagem. Os atributos exibidos correspondem ao número de vezes que cada rótulo de classe foi predito (por exemplo, #CN, #DA referem-se à quantidade de pontos CN e DA para um mesmo cérebro). A partir dos gráficos, observa-se a consistência e interpretabilidade do método. Nota-se que um cérebro CN tem mais pontos #CN do que pontos #DA e vice-versa. A mesma analogia pode ser estendida aos outros dois experimentos. Nota-se também, algumas sobreposições nos atributos das amostras para o experimento CN×CCL e uma sobreposição um pouco maior para o CCL×DA. Como os gráficos são para fins ilustrativos, essas sobreposições são perfeitamente aceitáveis, considerando que foram construídos usando apenas quatro dos dez atributos utilizados para a classificação. Além disso, visto que foram alcançados valores com alta acurácia (ou seja, 89,24%) para o experimento CN×DA, os atributos dessas amostras são mais visivelmente separados, mesmo quando usando apenas quatro atributos.

As Figuras 30, 31 e 32 ilustram exemplos visuais dos melhores e piores casos de cada experimento. Nesse caso, os melhores e piores casos foram definidos como aqueles que obtiveram a maior e a menor probabilidade de classificação (dentro os que acertaram a classificação), respectivamente.

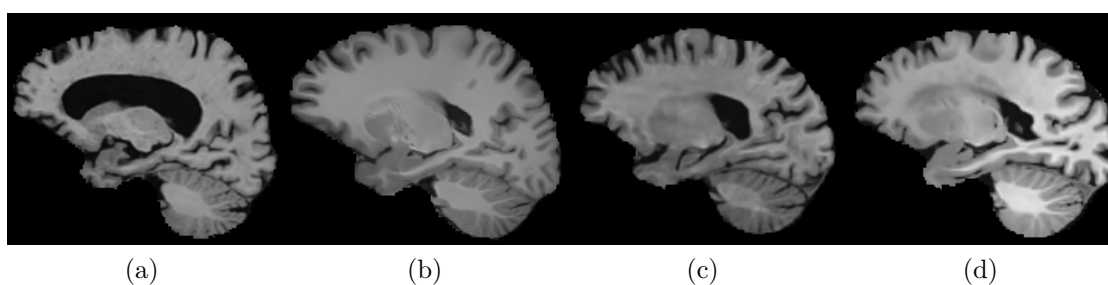


Figura 30 – Maiores e menores probabilidades de predição para o experimento CN×CCL: (a) Melhor para CCL, (b) melhor para CN, (c) pior para CCL e (d) pior para CN.

Para o experimento CN×CCL, as imagens que possuem maior probabilidade de predição apresentam padrões bem distintos, ou seja, a imagem com maior pro-

babilidade para a classe CCL, 90,78%, apresenta uma grande expansão do CSF e diminuição de GM, ao passo a que apresentou a maior probabilidade para a classe CN, 96,1%, é majoritariamente representada pelo tecido GM. Por outro lado, as imagens com menor probabilidade de predição, 61,08% para o CCL e 54,06% para o CN, visualmente apresentam quantidades similares dos tecidos.

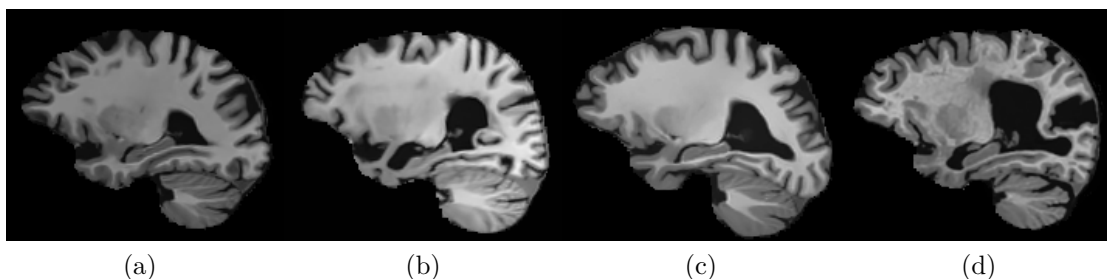


Figura 31 – Maiores e menores probabilidades de predição para o experimento CCL×DA: (a) Melhor para CCL, (b) melhor para DA, (c) pior para CCL e (d) pior para DA.

Para o experimento CCL×DA, o comportamento se repete, no entanto, com menores diferenças visuais entre as imagens que possuem maior probabilidade. As imagens CCL com maior probabilidade, 86,93%, apresentam uma pequena expansão do CSF em relação à diminuição de GM e as imagens DA com maior probabilidade, 85,11%, apresentam uma diferenciação um pouco maior entre os tecidos. As imagens com menor probabilidade de predição, 65,2% para o CCL e 51,26% para o DA, apresentam maior degeneração dos tecidos e a imagem com DA apresenta (visualmente) grande expansão do CSF em relação ao GM.

O mesmo padrão pode ser observado para o experimento CN×DA. As imagens CN e DA com maior probabilidade, de 87,48% e 97,77%, respectivamente, se diferenciam pela maior expansão do CSF na imagem com DA. As imagens com menor probabilidade de predição, 60,14% para a DA e 52,98% para o CN, apresentam diferenças mais sutis, com uma expansão do CSF maior para o CN do que para a DA.

Pelas análises das imagens foi possível observar que o método proposto conseguiu diferenciar imagens com padrões bastante semelhantes e até visualmente controversos, mostrando que o algoritmo pode extrair e diferenciar alterações es-

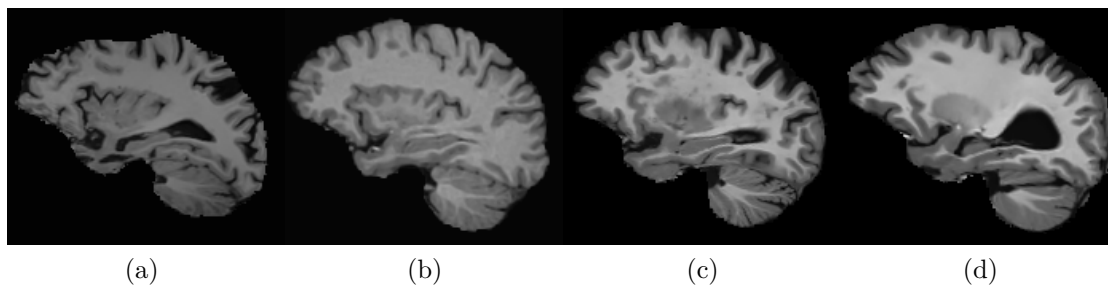


Figura 32 – Maiores e menores probabilidades de predição para o experimento CN×DA: (a) Melhor para DA, (b) melhor para CN, (c) pior para DA e (d) pior para CN.

truturais sutis entre as classes.

5.2 Análise de assimetrias estruturais

A metodologia utilizada nesta abordagem foi descrita na seção 4.5 e as informações sobre as imagens dos pacientes utilizados foram apresentadas na Tabela 5 da subseção 4.2.2. Os resultados para a classificação das imagens e a análise estatística dos atributos são apresentados nas subseções que seguem.

5.2.1 Configurações experimentais

Para a classificação das imagens de RM nos grupos de diagnóstico CN×CCL, CCL×DA, e CN×DA foram utilizados os atributos de correlação cruzada normalizada, calculados para cada medida estatística. Os hiperparâmetros dos classificadores foram estimados utilizando a técnica *nested cross-validation* com 10 *folds* e considerando a AUC como critério de desempenho. Foram testados os classificadores SVM com os *kernels* RBF, linear e polinomial. O melhor valor para o parâmetro C (em termos da maximização da AUC) de cada SVM foi determinado com a estratégia *coarse-grid* com potência de base 2, ou seja, 2^x , em que o expoente x varia entre $[-5, 10]$ com passo de 0,5; seguido de um *finer-grid*, realizado a partir do melhor valor encontrado no *coarse-grid*, \tilde{C} , com valores de $2^{\tilde{x}}$, em que o expoente \tilde{x} varia entre $[(\log_2 \tilde{C}) - 2, (\log_2 \tilde{C}) + 2]$ com passo de 0,25. Além disso, como pode ser observado na Tabela 5, os grupos de diagnóstico são levemente des-

balanceados, ou seja, existem quantidades distintas de imagens para cada grupo. Para contornar este problema, no projeto dos classificadores SVM foi realizado um ajuste nos pesos de cada grupo utilizando o parâmetro *class_weight*, mencionado anteriormente. Dessa maneira, os pesos de cada grupo foram ajustados automaticamente para serem inversamente proporcionais à quantidade de amostras de cada classe.

Por fim, os resultados de classificação médios foram calculados usando as métricas de acurácia, sensibilidade, especificidade, F1-score e AUC. Além disso, a classificação foi realizada utilizando individualmente cada uma das medidas estatísticas e com todas as combinações de cada medida. As métricas escolhidas para a exposição dos resultados foram, em sua maioria, calculadas a partir da matriz de confusão (vide seção 2.13). Nos experimentos CN×CCL e CN×DA, os grupos de diagnóstico CCL e DA foram considerados como positivos e o grupo CN como negativo, e no experimento CCL×DA, o grupo CCL foi considerado positivo e o DA negativo.

5.2.2 Análise dos parâmetros

Os parâmetros utilizados nas etapas propostas nesta abordagem foram avaliados a seguir.

5.2.2.1 Plano sagital médio

Para a filtragem das imagens e o cálculo da medida de *sheetness* (subseção 4.5.1) foi construído um banco de filtros de $64 \times 64 \times 64$ *voxels* usando os seguintes parâmetros: níveis de escala $N_s = 3$, ângulos azimutes $N_{a,0} = 3$, ângulos de elevação $N_e = 3$, frequência máxima central $\omega_{\max} = 0,2$, fator de escala $\gamma = 1,2$, largura de banda do filtro $\eta_\beta = 0,55$, e desvio angular $\sigma_\alpha = 1,2$. Como as imagens possuem dimensão $256 \times 256 \times 256$, todas foram reduzidas (*downsampling*) por um fator 4 para eficiência computacional. A forma dos filtros 3D log-Gabor foi definida visualmente com a ajuda de uma ferramenta Matlab, projetada (e gentilmente cedida ao nosso grupo) por Dosil, Pardo e Fdez-Vidal (2005), conforme ilustrado na Figura 33 e os parâmetros foram avaliados de forma visual visando ajustar a detecção do plano também para as imagens de pacientes com doenças neurodege-

nerativas, visto que os parâmetros sugeridos pelos autores foram ajustados para imagens de pacientes sem declínio cognitivo.

5.2.2.2 Filtragem direcional

O principal objetivo do projeto de banco de filtros foi criar filtros isotrópicos, dado que não há interesse em filtrar nenhuma orientação particular das estruturas dos hipocampos. Os valores dos parâmetros que melhor alcançaram a cobertura espectral uniforme usando filtros isotrópicos com sobreposição mínima foram os seguintes: fator de escala ($\gamma = 1,2$), largura de banda do filtro ($\eta_\beta = 0,55$) e desvio angular ($\sigma_\alpha = 1,2$).

Para definir o número de escalas (N_s), ângulos azimutes ($N_{a,0}$), ângulos de elevação (N_e) e frequência central máxima (ω_{\max}), foi realizado um experimento exploratório usando quarenta imagens de RM, sendo vinte CN e vinte AD não incluídas nos grupos de teste. Para isso, foi realizada uma análise usando todas as combinações dos seguintes valores de parâmetros $N_s = 2, 3$ e 4 , $N_{a,0} = 4, 6$ e 8 , $N_e = 3, 4$ e 5 e $\omega_{\max} = 0, 25, 0, 3, 0, 35$ e $0, 4$. Além disso, foi utilizada a função discriminante de Fisher (IZENMAN, 2013), calculada para cada medida estatística (média, variância, *skewness* e *kurtosis*), como figura de mérito a ser maximizada. Os melhores resultados foram obtidos para a medida da Variância com a seguinte combinação de parâmetros: $N_s = 3$, $N_{a,0} = 6$, $N_e = 4$ e $\omega_{\max} = 0, 25$, ou $0, 3$. Então, o valor de ω_{\max} foi variado no intervalo $[0, 25; -0, 3]$, com passos de $0, 01$ e o melhor resultado foi obtido para o parâmetro $\omega_{\max} = 0, 28$. Esse conjunto de parâmetros resultou em 23 orientações por escala sendo utilizado para a filtragem das imagens. A Figura 33 mostra uma representação visual do banco de filtros log-Gabor construído com esses parâmetros.

5.2.3 Resultados

Os resultados para a classificação das imagens e a análise estatística do IA desenvolvido são apresentadas a seguir.

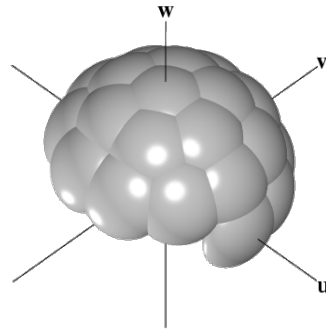


Figura 33 – Banco de filtros log-Gabor 3D utilizados nesta etapa.

5.2.3.1 Classificação

Para os experimentos de classificação foram utilizados os atributos obtidos da correlação cruzada normalizada, calculados para cada medida estatística, além da utilização de todas as possíveis combinações de cada medida estatística. É importante mencionar que nesta metodologia são utilizados vetores de atributos com baixa dimensionalidade, pois cada medida estatística possui um total de 26 atributos. Então, mesmo para o experimento que contém todas as concatenações das medidas estatísticas, o tamanho total é de apenas 104 atributos. Os melhores resultados obtidos para o treinamento das imagens foram utilizando o classificador SVM com *kernel* polinomial com grau unitário. Portanto, os resultados da classificação apresentados e discutidos são referentes a utilização desse *kernel*.

Os resultados médios de classificação com seus respectivos desvios-padrão, obtidos para cada uma das quatro medidas estatísticas, são apresentados na Tabela 12. Os melhores resultados em relação à AUC (para cada experimento) estão em negrito. Nos três experimentos de classificação, os maiores valores de AUC foram obtidos com a medida da Variância. Os valores de AUC e acurácia, respectivamente, para o experimento CN×CCL, foram iguais a 0,74 e 68,36%, enquanto para os casos CCL×DA e CN×DA, esses mesmos valores foram de 0,7 e 64,13% e 0,88 e 80,42%.

Após a realização dos experimentos de classificação usando as medidas estatísticas individualmente, foram realizados experimentos com duas, três e quatro combinações (concatenações horizontais) das medidas. Por simplicidade visual, os nomes das medidas estatísticas apresentadas a seguir foram abreviados; por exem-

Tabela 12 – Resultados da classificação das imagens utilizando *nested cross-validation* com 10 *folds* (média \pm desvio) para cada medida estatística.

		Média	Variância	<i>Kurtosis</i>	<i>Skewness</i>
CN×CCL	AUC	0,71 \pm 0,05	0,74 \pm 0,04	0,7 \pm 0,08	0,66 \pm 0,07
	Acc(%)	67,44 \pm 6,33	68,36 \pm 4,13	65,47 \pm 6,38	62,03 \pm 4,99
	Sens(%)	71,28 \pm 14,88	69,32 \pm 8,43	54,58 \pm 11,16	54,95 \pm 12,31
	Espec(%)	64,24 \pm 11,02	67,56 \pm 6,53	74,46 \pm 10,08	67,87 \pm 8,14
	F1	0,66 \pm 0,09	0,66 \pm 0,05	0,59 \pm 0,08	0,56 \pm 0,08
CCL×DA	AUC	0,62 \pm 0,07	0,7 \pm 0,12	0,61 \pm 0,08	0,66 \pm 0,09
	Acc(%)	58,91 \pm 7,97	64,13 \pm 11,19	60,0 \pm 7,75	61,3 \pm 7,79
	Sens(%)	57,42 \pm 10,22	59,45 \pm 16,91	64,97 \pm 13,85	63,35 \pm 8,96
	Espec(%)	60,79 \pm 9,64	69,76 \pm 9,53	54,02 \pm 10,07	58,81 \pm 8,7
	F1	0,6 \pm 0,08	0,64 \pm 0,13	0,63 \pm 0,09	0,64 \pm 0,08
CN×DA	AUC	0,82 \pm 0,03	0,88 \pm 0,03	0,79 \pm 0,07	0,79 \pm 0,06
	Acc(%)	73,77 \pm 4,83	80,42 \pm 5,73	76,33 \pm 6,83	74,36 \pm 6,13
	Sens(%)	74,17 \pm 5,08	81,79 \pm 6,4	69,83 \pm 11,3	70,24 \pm 10,72
	Espec(%)	73,48 \pm 5,54	79,48 \pm 7,73	80,82 \pm 6,34	77,15 \pm 5,75
	F1	0,7 \pm 0,05	0,77 \pm 0,06	0,7 \pm 0,09	0,69 \pm 0,08

plo, *Var&Kurt* corresponde à concatenação dos vetores de atributos da Variância e da *Kurtosis*.

A Tabela 13 apresenta os resultados médios da classificação obtidos para a concatenação de duas medidas estatísticas. Os melhores resultados (em relação a AUC) foram alcançados usando os vetores de *Média&Var*, *Var&Kurt* e *Var&Skew*, respectivamente, para os experimentos CN×CCL, CCL×DA e CN×DA. Os valores de AUC e respectivas acurácias foram de 0,76 e 65,65% para CN×CCL, 0,7 e 65,65% para CCL×DA e 0,9 e 81,42% para CN×DA. De maneira geral, ao compararmos os resultados de AUC apresentados nas Tabelas 12 e 13, nota-se que os resultados foram positivamente influenciados pela concatenação dos vetores de atributos.

A Tabela 14 apresenta os resultados médios da classificação obtidos para a concatenação de três e quatro medidas estatísticas. Os melhores resultados (em relação a AUC) foram obtidos para os vetores de *Média&Var&Kurt*, *Var&Kurt&Skew* e *Média&Var&Kurt*, com valores de AUC e respectivas acurácias, obtidos para os experimentos CN×CCL, CCL×DA e CN×DA, de 0,76 e 69,44%, 0,69 e 64,57% e 0,9 e 82,59%, respectivamente. Mais uma vez, é possível observar que os resultados da AUC foram positivamente influenciados pela concatenação dos vetores de atributos, embora o uso de todos os quatro não tenha atingido os melhores valores.

Tabela 13 – Resultados da classificação das imagens utilizando *nested cross-validation* com 10 *folds* (média \pm desvio) para a concatenação de duas medidas estatísticas

		Média&Var	Média&Kurt	Média&Skew	Var&Kurt	Var&Skew	Kurt&Skew
CN×CCL	AUC	0,76 \pm 0,05	0,74 \pm 0,07	0,73 \pm 0,08	0,75 \pm 0,07	0,75 \pm 0,06	0,7 \pm 0,08
	Acc(%)	65,65 \pm 4,71	68,53 \pm 5,87	64,74 \pm 8,16	66,53 \pm 6,78	68,19 \pm 5,35	65,47 \pm 6,25
	Sens(%)	65,68 \pm 8,41	66,54 \pm 12,93	64,88 \pm 15,95	64,95 \pm 9,5	64,52 \pm 13,7	58,55 \pm 8,93
	Espec(%)	65,61 \pm 7,35	70,17 \pm 6,15	64,63 \pm 11,68	67,82 \pm 10,45	71,23 \pm 8,54	71,19 \pm 9,15
	F1	0,63 \pm 0,06	0,65 \pm 0,08	0,62 \pm 0,1	0,64 \pm 0,07	0,64 \pm 0,08	0,61 \pm 0,07
CCL×DA	AUC	0,69 \pm 0,1	0,64 \pm 0,09	0,65 \pm 0,1	0,7 \pm 0,09	0,69 \pm 0,09	0,63 \pm 0,09
	Acc(%)	62,17 \pm 7,11	60,87 \pm 7,25	58,7 \pm 9,05	65,65 \pm 7,79	65,43 \pm 7,97	60,43 \pm 8,62
	Sens(%)	62,17 \pm 6,45	61,37 \pm 9,72	55,78 \pm 10,85	63,35 \pm 9,9	62,18 \pm 8,88	66,92 \pm 11,02
	Espec(%)	62,17 \pm 11,01	60,26 \pm 11,97	62,12 \pm 9,06	68,4 \pm 8,81	69,36 \pm 8,83	52,6 \pm 6,81
	F1	0,64 \pm 0,06	0,63 \pm 0,07	0,59 \pm 0,1	0,67 \pm 0,08	0,66 \pm 0,08	0,65 \pm 0,08
CN×DA	AUC	0,89 \pm 0,04	0,85 \pm 0,04	0,84 \pm 0,04	0,89 \pm 0,04	0,9 \pm 0,04	0,82 \pm 0,07
	Acc(%)	80,42 \pm 5,72	78,09 \pm 5,4	74,97 \pm 6,02	81,04 \pm 7,17	81,42 \pm 6,3	76,35 \pm 6,98
	Sens(%)	79,88 \pm 7,79	75,07 \pm 7,93	72,21 \pm 9,06	78,98 \pm 7,78	79,0 \pm 8,39	72,69 \pm 9,67
	Espec(%)	80,82 \pm 9,06	80,12 \pm 6,25	76,85 \pm 6,05	82,51 \pm 9,49	83,14 \pm 7,82	78,86 \pm 6,57
	F1	0,77 \pm 0,06	0,74 \pm 0,06	0,7 \pm 0,07	0,77 \pm 0,08	0,78 \pm 0,07	0,71 \pm 0,08

Tabela 14 – Resultados da classificação das imagens utilizando *nested cross-validation* com 10 *folds* (média \pm desvio) para a concatenação de três e quatro medidas estatísticas

		Média&Var&Kurt	Média&Var&Skew	Var&Kurt&Skew	Média&Var&Kurt&Skew
CN×CCL	AUC	0,76 \pm 0,07	0,74 \pm 0,08	0,74 \pm 0,05	0,75 \pm 0,07
	Acc(%)	69,44 \pm 4,83	66,74 \pm 4,61	68,0 \pm 4,73	69,27 \pm 4,54
	Sens(%)	68,09 \pm 12,86	65,69 \pm 10,3	64,51 \pm 12,51	66,89 \pm 11,09
	Espec(%)	70,53 \pm 9,19	67,56 \pm 8,55	70,84 \pm 9,3	71,18 \pm 9,92
	F1	0,66 \pm 0,07	0,64 \pm 0,06	0,64 \pm 0,07	0,66 \pm 0,06
CCL×DA	AUC	0,32 \pm 0,09	0,69 \pm 0,1	0,69 \pm 0,08	0,69 \pm 0,07
	Acc(%)	38,91 \pm 7,49	61,52 \pm 9,23	64,57 \pm 8,01	63,7 \pm 5,98
	Sens(%)	41,43 \pm 12,65	59,38 \pm 11,48	63,8 \pm 11,72	60,57 \pm 10,51
	Espec(%)	35,95 \pm 11,14	64,05 \pm 9,15	65,5 \pm 6,62	67,45 \pm 4,98
	F1	0,42 \pm 0,1	0,62 \pm 0,09	0,66 \pm 0,09	0,64 \pm 0,07
CN×DA	AUC	0,9 \pm 0,04	0,9 \pm 0,04	0,88 \pm 0,05	0,89 \pm 0,05
	Acc(%)	82,59 \pm 6,84	80,24 \pm 6,47	81,23 \pm 5,26	81,23 \pm 7,72
	Sens(%)	74,62 \pm 9,6	76,57 \pm 8,18	78,48 \pm 7,5	78,45 \pm 7,58
	Espec(%)	88,1 \pm 7,67	82,78 \pm 7,81	83,14 \pm 6,24	83,17 \pm 10,86
	F1	0,78 \pm 0,09	0,76 \pm 0,07	0,77 \pm 0,06	0,78 \pm 0,08

As Figuras 34(a), 34(b) e 34(c) mostram as curvas ROC e seus valores da AUC correspondentes às classificações CN×CCL, CCL×DA e CN×DA, respectivamente. As curvas ROC são referentes aos melhores resultados apresentados em cada uma das Tabelas 12, 13 e 14. A partir das curvas é possível observar que, embora as concatenações tenham influenciado positivamente os resultados, a forma

das curvas ROC e, conseqüentemente, suas áreas (AUCs) são bastante semelhantes. Vale também destacar que, embora os valores de acurácia apresentados tenham sido ligeiramente inferiores aos da maioria dos resultados relatados na literatura, os valores das AUCs foram altos. Esses resultados podem ser explicados pelo fato dos experimentos terem sido realizados usando um conjunto de dados levemente desbalanceado. Ressalta-se ainda que, diferentemente da acurácia, a AUC é menos sensível a esse desbalanceamento de classes (NARASIMHAMURTHY, 2017).

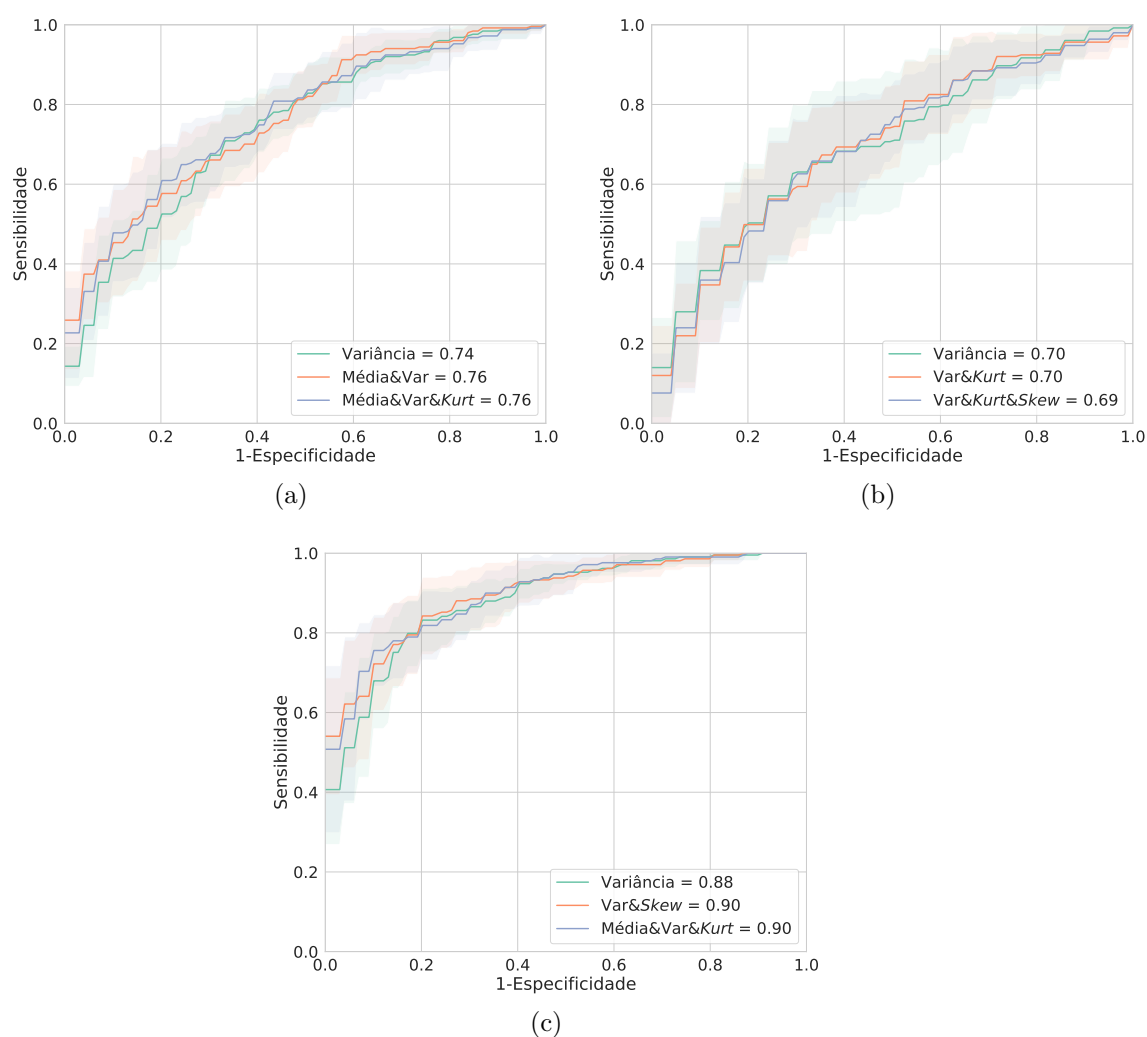


Figura 34 – Curvas ROC e valores de AUC correspondentes aos resultados de classificação de (a) $CN \times CCL$, (b) $CCL \times DA$ e (c) $CN \times DA$.

Um ponto importante a ser destacado ao analisar conjuntamente os resultados das Tabelas 12, 13 e 14 é que as concatenações contendo o atributo Variância sempre proporcionaram ganhos maiores, o que pode ser um indicativo de que essa métrica conseguiu extrair atributos com maiores diferenças hemisféricas entre as classes que as demais. Também é possível observar que, para quase todos os experimentos, os valores de especificidade foram levemente maiores que os valores de sensibilidade, indicando que o classificador teve maiores acertos nos grupos negativos do que nos positivos, o que poderia ser problemático, visto que nos experimentos as classes positivas são as portadoras dos diagnósticos mais importantes. No entanto, como esse trabalho utiliza curvas ROC (Figuras 34(a), 34(b) e 34(c)), um ponto de operação desejado, que se traduz na relação entre os valores de sensibilidade e especificidade, pode ser escolhido para ser utilizado pelos especialistas do domínio. Por fim, pode-se dizer que os resultados obtidos da abordagem proposta foram muito promissores, principalmente porque os experimentos de classificação usaram apenas atributos assimétricos extraídos dos hipocampos.

5.2.3.2 Análise estatística

Para a análise estatística foram utilizados os quatro *IAs* provindos da correlação cruzada normalizada, $NCC(x)$, de cada vetor de medida estatística. As análises foram realizadas utilizando a biblioteca *statsmodels*⁴ da linguagem *python*. Também foram testadas as diferenças estatísticas entre os valores de *IA* em relação aos grupos de diagnóstico usando os testes ANOVA e HSD, ambos com $\alpha = 0,05$. A análise para as diferenças entre gêneros foi realizada apenas com o teste ANOVA com $\alpha = 0,05$.

IA vs. grupos de diagnóstico

Os resultados da ANOVA para os quatro *IAs* comparados aos grupos de diagnósticos, seus respectivos p-valores e as conclusões obtidas dos testes *post-hoc* são apresentados na Tabela 15. É possível verificar que as três medidas estatísticas (Média, Variância e *Kurtosis*) demonstraram diferenças estatisticamente significativas (p-valor menor que 0,05) entre as médias dos grupos de diagnóstico. A

⁴ <<https://pypi.org/project/statsmodels/>>

medida da *Skewness*, por apresentar um p-valor maior que 0,05, não demonstrou diferença estatisticamente significativa entre as médias dos grupos de diagnóstico. Sendo assim, somente as medidas que passaram no ANOVA (Média, Variância e *Kurtosis*) foram utilizadas para a execução do teste HSD. Nesse caso, foram realizados testes pareados entre os grupos de diagnóstico, sendo CN×CCL, CCL×DA e CN×DA. Segundo as conclusões dos testes, os pares CN×CCL e CN×DA apresentaram diferenças estatisticamente significativas para todas as medidas estatísticas comparadas. O par CCL×DA mostrou diferenças significantes apenas para a medida de Variância.

Tabela 15 – Média e desvio padrão dos valores de *IA* por grupos de diagnóstico, resultados da ANOVA ($\alpha = 0,05$) e conclusões dos testes *post-hoc* HSD ($\alpha_t < 0,05$).

Índice de Assimetria	CN	CCL	DA	ANOVA	
				p-valor	<i>post-hoc</i> HSD
Média	0,6332 ± 0,0099	0,6389 ± 0,0073	0,6382 ± 0,0075	< 0,001	CN<CCL, CN<DA
Variância	0,7120 ± 0,0083	0,7162 ± 0,0088	0,7182 ± 0,0095	< 0,001	CN<CCL, CCL<DA, CN<DA
<i>Kurtosis</i>	0,6776 ± 0,0716	0,7099 ± 0,0803	0,7165 ± 0,0869	< 0,001	CN<CCL, CN<DA
<i>Skewness</i>	0,3198 ± 0,0313	0,3235 ± 0,0321	0,3209 ± 0,0315	0,399	Não significativo.

Outro ponto a ser observado é que as medidas estatísticas que passaram no teste ANOVA, e obtiveram conclusões no teste HSD, apresentaram uma tendência de assimetria, ou seja, menores valores de *IA* a medida em que a condição neurodegenerativa se agrava, (CN<CCL<DA). Essa tendência, relatada em outros trabalhos presentes na literatura de neurologia (SARICA et al., 2018; ARDEKANI et al., 2019; SHI et al., 2009), mostra um aumento na assimetria dos hemisférios cerebrais à medida que a gravidade da doença de Alzheimer aumenta. A Figura 35(a) mostra o *boxplot* do *IA* da variância para a população estratificada por grupos de diagnóstico. Ao visualizar o *boxplot*, nota-se que os valores de *IA* aumentaram à medida que a gravidade do declínio cognitivo aumentou.

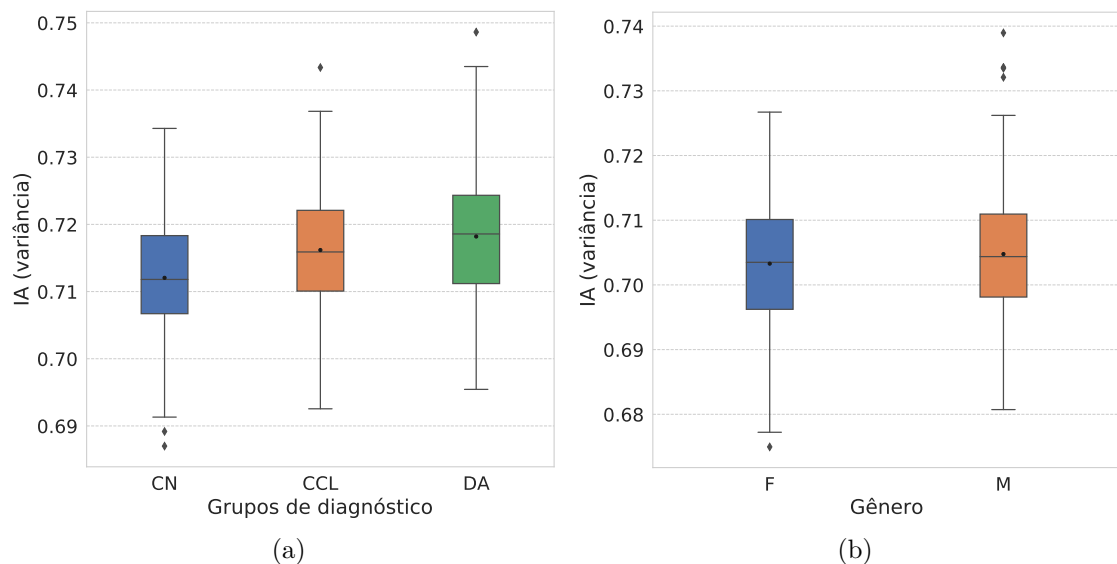
IA vs. gênero

A Tabela 15 apresenta os resultados da ANOVA para os *IAs* comparados aos gêneros e seus respectivos p-valor. Como existem apenas duas médias para serem comparadas, Feminina (F) e Masculina (M), o teste HSD não foi executado.

Tabela 16 – Média e desvio padrão dos valores de IA por gênero e resultados da ANOVA ($\alpha = 0,05$).

Índice de Assimetria		F	M	ANOVA p-valor
Média	$0,6209 \pm 0,0105$	$0,6236 \pm 0,0079$		$< 0,001$
Variância	$0,7033 \pm 0,0093$	$0,7048 \pm 0,0096$		$0,033$
<i>Kurtosis</i>	$0,6707 \pm 0,0763$	$0,6997 \pm 0,0870$		$< 0,001$
<i>Skewness</i>	$0,2962 \pm 0,0331$	$0,2945 \pm 0,0326$		$0,497$

Nessa comparação, assim como na anterior, é possível verificar que, por apresentaram p-valor menor que 0,05, a Média, a Variância e a *Kurtosis* demonstraram diferenças estatisticamente significativas entre as médias dos gêneros e a medida da *Skewness* não demonstrou diferença estatisticamente significativa (p-valor maior que 0,05). Além disso, considerando apenas as métricas que passaram no teste ANOVA, é possível verificar que os valores de IA foram ligeiramente mais baixos para o grupo feminino, uma tendência que também foi relatada em outros trabalhos da literatura em neurologia (ARDEKANI et al., 2019; LUCARELLI et al., 2013). A Figura 35(b) mostra o *boxplot* da medida de variância para a população estratificada por gênero.

Figura 35 – Boxplots das diferenças entre as médias de IA da variância por (a) grupos de diagnóstico e (b) gênero.

Para as duas análises estatísticas realizadas, a medida da *Skewness* não apresentou diferenças estatisticamente significantes nem entre os grupos de diagnóstico e nem entre os gêneros. Além disso, a única medida estatística que obteve conclusões para o par CCL×DA foi a Variância. Apesar de não ser uma comparação direta, se relacionarmos as conclusões estatísticas com os resultados de classificação, os experimentos de classificação que continham o atributo Variância tiveram os melhores resultados, possivelmente por essa medida estatística possuir diferença estatisticamente significativa entre todos os grupos de diagnóstico comparados.

5.3 Análise do erro de predição da idade biológica do cérebro

A metodologia utilizada nesta abordagem foi descrita na seção 4.6 com as imagens de treinamento descritas na subseção 4.6.1 e os aumentos de dados descritos na subseção 4.6.4. As configurações experimentais e os resultados são apresentados a seguir.

5.3.1 Configurações experimentais

Conforme mencionado na seção 4.6, foram realizados dois experimentos. Para cada um, foram treinadas duas redes diferentes para o hipocampo direito e esquerdo, e os resultados das predições foram combinados (*ensemble*) para criar a predição final da rede. Os resultados mostraram que o uso do *ensemble* aumentou a estabilidade das predições da rede, com redução da variância, devido ao uso de dados mais complementares, ou seja, duas ROIs diferentes do mesmo cérebro.

Para avaliação do desempenho, foram calculadas três métricas, o MAE, RMSE e o $\Delta_{\text{Idade cérebro}}$, para os resultados de predição de cada hipocampo e do *ensemble*. Além disso, o coeficiente de correlação de Pearson (r) foi calculado para o primeiro treinamento.

Para encontrar o melhor conjunto de hiperparâmetros, o conjunto de dados de validação foi avaliado utilizando um procedimento de validação cruzada estratificada com dois *folds*, usando o MAE como função de perda. Para estratificar as idades dos pacientes foi realizada uma discretização considerando um intervalo de

três anos. A Figura 36 mostra a distribuição de idades para os grupos de validação e teste para ambos os experimentos. A partir dos histogramas observa-se que os dados de teste e validação possuem distribuições semelhantes.

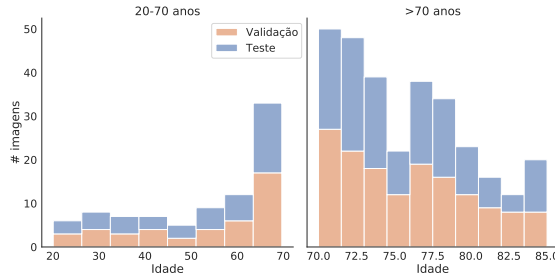


Figura 36 – A distribuição de idade da população para as imagens de validação e teste.

Todos os modelos foram treinados usando a estratégia de taxa de aprendizagem CLR (Smith, 2017) (subseção 2.12.2) com decaimento exponencial. Para determinar os limites ideais para a inicialização CLR, o modelo foi treinado por 50 épocas com um aumento linear da TA entre $1e^{-10}$ e $1e^{+1}$. A seguir, foi realizada uma análise dos valores da TA em conjunto com a função de perda ao longo das épocas para encontrar o “intervalo ideal”, conforme sugerido em (Smith, 2017). A Figura 37 ilustra o experimento realizado.

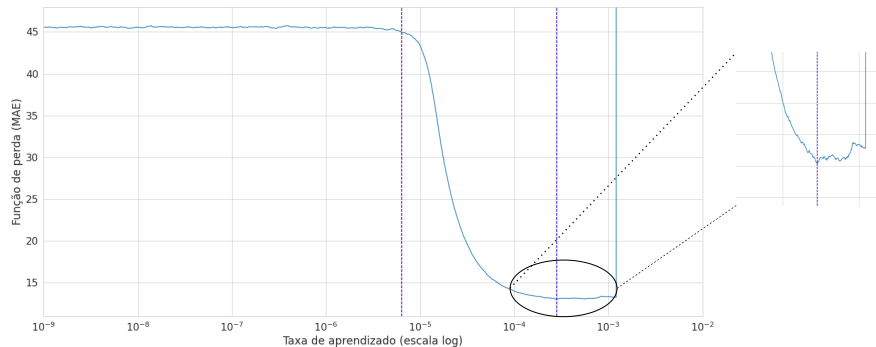


Figura 37 – Análise da faixa ótima da TA.

Conforme o gráfico apresentado, é possível observar que a função de perda não diminui para valores baixos de TA. A faixa ótima corresponde aos valores entre $10e^{-5,2}$ (*base_lr*) e $10e^{-3,6}$ (*max_lr*), em que se nota um valor associado a uma

queda mais acentuada da função de perda, seguida por um pequeno aumento. A seguir, observa-se também que para valores maiores do que o valor de max_lr a função de perda aumenta à medida que o parâmetro é atualizado, causando um salto e até mesmo na divergência dos valores mínimos.

Diferentes técnicas de otimização e regularização foram testadas usando cada um dos respectivos limites da CLR encontrados. Para a otimização foram testados os algoritmos SGD, SGD com *momentum* (QIAN, 1999), Adam (KINGMA; BA, 2014), e RMSprop (TIELEMAN; HINTON, 2012). Para evitar *overfitting* foram testadas as seguintes técnicas de regularização: *weight decay* (L2), *dropout* e aumento de dados. O primeiro conjunto de experimentos, contendo pacientes com faixa etária entre 20 e 70 anos, foi avaliado por 150 épocas e o segundo, contendo pacientes com faixa etária maior que 70 anos, por 100 épocas.

Todos os experimentos foram implementados usando o PyTorch (v. 1.7.1) e os modelos foram treinados em duas GPUs NVIDIA 1080-TI com paralelização dos *batches*. Os hiperparâmetros usados para treinar as redes neurais são mostrados na Tabela 17. O código-fonte estará disponível no GitHub⁵ após a publicação.

Tabela 17 – Hiperparâmetros usados nos experimentos.

Hiperparâmetros							
Faixa etária	hip.	CLR (<i>exp_range</i>)	taxa de <i>dropout</i>	tam. do <i>batch</i>	otimizador	# Max de épocas	Melhor época
20-70	esquerdo	$[10e^{-5.2}, 10e^{-3.4}]$	0.3	128	RMSprop	150	117
	direito	$[10e^{-5.2}, 10e^{-3.6}]$	0.2	128	RMSprop	150	142
>70	esquerdo	$[10e^{-4.7}, 10e^{-3.3}]$	0.2	128	RMSprop	100	83
	direito	$[10e^{-4.7}, 10e^{-3.3}]$	0.2	128	RMSprop	100	95

5.3.2 Resultados

Os resultados dos experimentos para as regiões hipocampal esquerda e direita e *ensemble* dos dois treinamentos, que corresponde à média das duas predições da rede, são apresentados nas subseções a seguir.

5.3.2.1 Experimento 1: pacientes com faixa etária entre 20 e 70 anos

A análise dos modelos foi iniciada com a investigação da influência das técnicas de regularização, por exemplo, aumento de dados, taxas de *dropout* e *weight decay*.

⁵ https://github.com/kapoloni/age_prediction

Para isso, as curvas da função de perda dos modelos foram plotadas e os padrões entre o treinamento e a validação avaliados. A Figura 38 mostra as curvas da função de perda contrastando o estágio de treinamento com e sem aumento de dados por 150 épocas. Pelo gráfico, pode-se notar que o modelo sendo treinado sem aumento de dados teve problemas com mínimos locais ou pontos de sela e não melhorou ao longo das épocas. Em contraste, o modelo com aumento dos dados melhorou significativamente os resultados do treinamento.

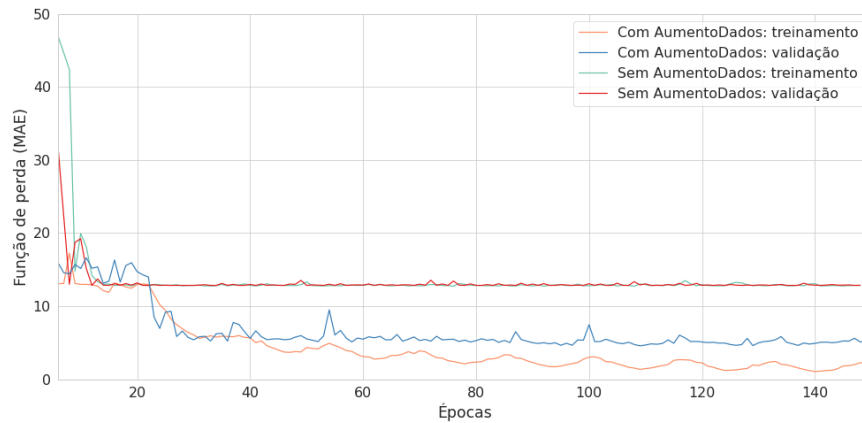


Figura 38 – Curvas da função de perda do modelo com e sem aumento de dados.

Além disso, as taxas de *dropout* e a influência do *weight decay* foram avaliadas usando o aumento de dados; no entanto, os modelos tiveram problemas com mínimos locais com o uso de *weight decay* e altas porcentagens de *dropout* (o *dropout* foi aplicado somente nos neurônios que antecedem a última camada totalmente conectada da rede). Em relação à avaliação dos otimizadores, o RMSprop, o mesmo otimizador usado no artigo *EfficientNet* (TAN; LE, 2019), obteve o menor valor para a função de perda em todos os experimentos. Portanto, os resultados são apresentados e discutidos usando os hiperparâmetros mencionados na Tabela 17. Em suma, foi utilizado aumento de dados, taxas de *dropout* de 0,2 e 0,3, um tamanho de *batch* de 128 (64 considerando a paralelização) e o otimizador RMSprop. Este primeiro conjunto de experimentos foi executado por 150 épocas e o modelo selecionado foi aquele que obteve o menor valor médio para a função de perda no conjunto de validação ao longo das épocas.

A Tabela 18 mostra os resultados do primeiro treinamento. Nota-se que os

resultados do hipocampo direito e esquerdo são muito semelhantes e que o *ensemble* melhorou ainda mais os resultados. Os melhores resultados para o conjunto de teste, obtidos para o *ensemble* dos modelos, foram MAE de 3,31, RMSE de 4,65, r de 0,95 e um $\Delta_{\text{Idade cérebro}}$ de $-0,68$.

Tabela 18 – Resultados da predição da idade do cérebro para o primeiro experimento.

Hip.	Validação				Teste			
	MAE	RMSE	$\Delta_{\text{Idade cérebro}}$	r	MAE	RMSE	$\Delta_{\text{Idade cérebro}}$	r
Esquerdo	4,8	6,29	0,49	0,92	4,24	6,03	0,17	0,93
Direito	4,24	5,81	-0,04	0,92	4,71	6,11	-1,54	0,93
<i>Ensemble</i>	3,87	5,01	0,23	0,95	3,31	4,65	-0,68	0,95

A Figura 39 mostra os gráficos de correlação das idades biológicas do cérebro estimadas com as idades cronológicas para cada região do hipocampo e *ensemble*. Analisando os gráficos observa-se que ambas as redes alcançaram diferentes predições de idade individual, apesar dos resultados da correlação (r) para o hipocampo direito e esquerdo serem iguais, 0,93. As diferenças podem ser notadas avaliando os gráficos dos histogramas marginais. Os bins das idades estimadas e das idades cronológicas são distribuídos de forma semelhante para o hipocampo esquerdo, mas mostram uma diferença maior entre bins correspondentes à segunda maior categoria de idade, ou seja, aproximadamente entre 55 e 63 anos. O mesmo padrão pode ser encontrado para o hipocampo direito, mas com uma diferença maior entre os bins correspondentes às idades menores, aproximadamente entre 20 e 27 anos. Essas diferenças entre as predições dos modelos podem ser complementares e ajudar a melhorar os resultados. Ao combinar as predições (modelo *ensemble*), obteve-se um valor de correlação superior, $r = 0,95$, e uma distribuição do histograma muito semelhante entre as idades estimadas e as cronológicas. Além disso, foi calculada a correlação de Pearson e um valor de $p < 0,001$ foi obtido para os três modelos, o que implica que as idades estimadas e as idades cronológicas apresentam uma relação linear estaticamente significativa.

Para verificar se os erros de predição do modelo criado possuem explicações biológicas ou foram causadas por falha do modelo, as imagens hipocampais de dois pacientes com a mesma idade apresentando valores baixo e alto de $\Delta_{\text{Idade cérebro}}$

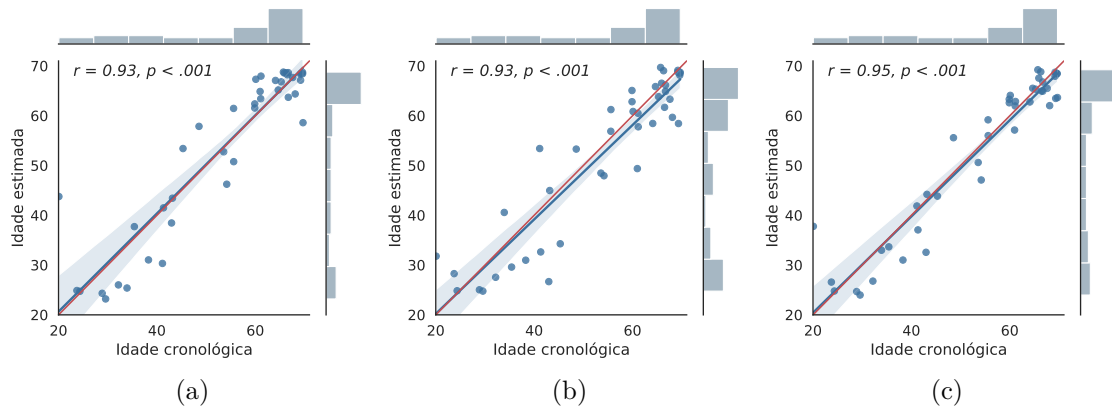


Figura 39 – Gráficos de correlação das idades biológicas estimadas com as idades cronológicas do modelo CNN para (a) o hipocampo esquerdo, (b) o hipocampo direito e (c) o *ensemble*.

foram comparadas. A Figura 40 mostra as imagens das vistas axiais, coronais, e sagital dos dois pacientes. Os resultados de ambos os pacientes podem ser identificados ao examinar os gráficos de correlação na Figura 39. Os pacientes 1 e 2 tinham aproximadamente 43 anos de idade, entretanto a idade biológica predita do cérebro do paciente 1 foi de 32, 56 anos, enquanto que para o paciente 2 foi de 44, 2. Assim, o paciente 1 apresentou um $\Delta_{\text{Idade cérebro}}$ de $-10, 41$, o que significa que o modelo subestimou a idade do cérebro em pelo menos dez anos. No caso do paciente 2, o valor do $\Delta_{\text{Idade cérebro}}$ foi de $1, 04$, significando que o modelo superestimou a idade do cérebro em cerca de um ano. Além disso, ao analisar as imagens, observa-se que o cérebro do paciente 1 tem uma aparência de um cérebro mais jovem, ou seja, apresenta menor atrofia do que o paciente 2, o que pode ser evidenciado pela baixa presença de tecido do CSF (cor mais escura) nas imagens do paciente 1. Essas diferenças podem ser explicadas pela extensa gama de variações “saúáveis” existentes no cérebro humano com o envelhecimento, com mudanças anatômicas internas ao paciente e entre uma população (DINSDALE et al., 2021). Portanto, essas diferenças criam dificuldades para a predição do modelo, diminuindo, assim, a sua precisão.

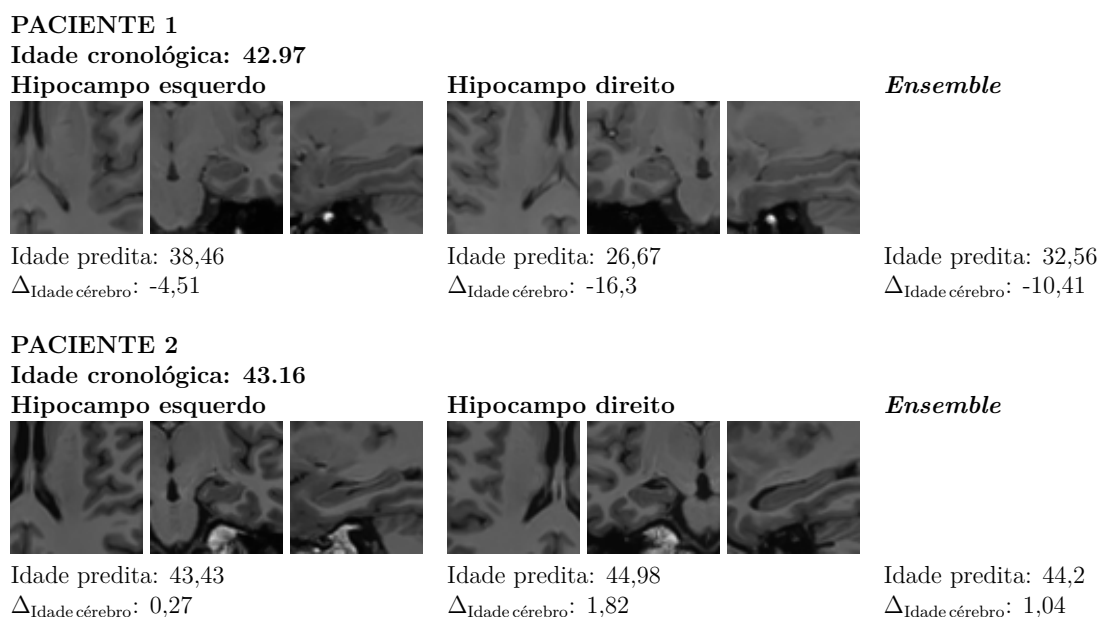


Figura 40 – Imagens de pacientes com idade cronológica de ≈ 43 anos.

5.3.2.2 Experimento 2: pacientes com faixa etária maior que 70 anos

Para o segundo conjunto de experimentos, foram utilizados os pesos pré-treinados dos melhores modelos do primeiro experimento e realizado um *fine-tuning*. Como os pesos não são aleatórios, os experimentos foram executados por apenas 100 épocas. Os modelos selecionados correspondem aos que obtiveram a menor função de perda no conjunto de validação. Assim como nos experimentos anteriores, a influência das técnicas de regularização e otimizadores também foram avaliados. O melhor conjunto de hiperparâmetros foi apresentado na Tabela 17. Em suma, foi utilizado aumento de dados, taxa de *dropout* de 0,2, tamanho de *batch* de 128 e o otimizador RMSprop.

A Tabela 19 mostra os resultados de validação e teste do segundo treinamento. Os resultados do hipocampo direito e esquerdo são muito semelhantes e o *ensemble* melhorou os resultados. Nesse caso, os melhores resultados obtidos para o conjunto de teste foram: MAE de 3,66, RMSE de 4,58 e um $\Delta_{\text{Idade cérebro}}$ de 1,24. Conforme esperado, os resultados são bastante semelhantes aos do primeiro experimento, visto que foi realizado apenas um *fine-tuning* dos parâmetros do modelo para prever a idade de pacientes CN idosos.

Tabela 19 – Resultados da predição da idade do cérebro para o segundo experimento.

Hip.	Validação			Teste		
	MAE	RMSE	$\Delta_{\text{Idade cérebro}}$	MAE	RMSE	$\Delta_{\text{Idade cérebro}}$
Esquerdo	3,74	4,7	0,36	4,05	5,12	0,82
Direito	3,85	4,92	1,29	3,96	4,98	1,65
<i>Ensemble</i>	3,52	4,39	0,82	3,66	4,58	1,24

A Tabela 20 apresenta os resultados do segundo treinamento, avaliado com imagens pertencentes aos pacientes com CCL e DA. Comparando com os resultados dos testes anteriores (imagens de pacientes CN), nota-se que ocorreu uma superestimação do $\Delta_{\text{Idade cérebro}}$ para os grupos CCL e DA, seguindo o seguinte padrão: $\text{CN} < \text{CCL} < \text{DA}$. Além disso, também é possível notar esse mesmo padrão para as métricas MAE e RMSE.

Tabela 20 – Resultados da predição da idade do cérebro para os grupos com CCL e DA.

Hip.	CCL			DA		
	MAE	RMSE	$\Delta_{\text{Idade cérebro}}$	MAE	RMSE	$\Delta_{\text{Idade cérebro}}$
Esquerdo	4,79	6,04	2,21	5,48	6,71	3,88
Direito	6,38	7,56	4,6	6,82	8,22	6,13
<i>Ensemble</i>	5,19	4,45	3,4	5,8	7,05	5,01

Análise estatística

Para verificar se o padrão apresentado é estatisticamente significativo, foram realizados os testes ANOVA e Tukey HSD. A Tabela 21 apresenta os resultados desses testes, que incluem a média e o desvio padrão dos valores do $\Delta_{\text{Idade cérebro}}$, o p-valor da ANOVA e as conclusões após a execução do *post-hoc* HSD.

Os valores apresentados possuem uma tendência de aumento conforme a gravidade do diagnóstico: os valores são mais baixos no grupo CN do que no CCL e no CCL comparado ao DA. Na execução do teste *post-hoc* entre os pares de grupos de diagnóstico, todos mostraram diferenças significantes. Essa tendência é consis-

Tabela 21 – Média e desvio padrão dos valores do $\Delta_{\text{Idade cérebro}}$ por grupos de diagnóstico, resultados da ANOVA ($\alpha = 0,05$) e conclusões dos testes *post-hoc* HSD ($\alpha_t < 0,05$).

	CN	CCL	DA	ANOVA	
				<i>p</i> -valor	<i>post-hoc</i>
$\Delta_{\text{Idade cérebro}}$	1,24 ± 4,42	3,4 ± 5,32	5,01 ± 4,97	< 0,001	CN<CCL, CN<DA, CCL<DA

tente, pois mostra um aumento no envelhecimento biológico do cérebro à medida que a gravidade do Alzheimer aumenta. Conforme já mencionado, o cérebro sofre alterações estruturais mais intensas e acentuadas durante o desenvolvimento da DA. Portanto, a superestimação das idades biológicas dos cérebros dos pacientes com DA e CCL é esperada. A Figura 41 mostra o gráfico de *boxplot* dos valores $\Delta_{\text{Idade cérebro}}$ para cada grupo de diagnóstico.

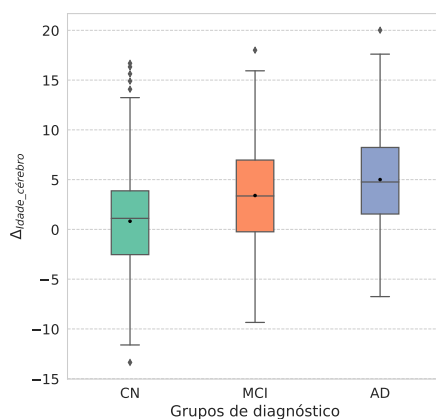


Figura 41 – *Boxplots* dos valores $\Delta_{\text{Idade cérebro}}$ por grupo de diagnóstico.

Por fim, foi calculada a correlação de Pearson para medir a força e a direção da associação entre o MMSE (teste clínico — seção 2.1) e os valores $\Delta_{\text{Idade cérebro}}$ obtidos, conforme ilustrado na Figura 42. Nesse caso, foi encontrada uma correlação estatisticamente significativa entre as variáveis (*p*-valor < 0,001) e, conforme o esperado, foi obtido um coeficiente de correlação negativo, $r = -0,31$, visto que ambos os valores possuem uma relação inversa, ou seja, valores menores de MMSE e maiores de $\Delta_{\text{Idade cérebro}}$ indicam maior gravidade da doença. Embora o valor do coeficiente de correlação não tenha sido alto nesta comparação, nota-se uma relação estreita entre o envelhecimento cerebral patológico e a piora potencial

do funcionamento cognitivo. Conforme já relatado em outros estudos (FRANKE et al., 2010; GASER et al., 2013), pontuações delta mais altas estão intimamente relacionadas às medidas de gravidade da doença clínica em pacientes com DA. Esses resultados apoiam fortemente a relação entre o envelhecimento cerebral profundamente acelerado e a gravidade da doença, mais pronunciada em indivíduos já diagnosticados com DA, e a piora potencial do funcionamento cognitivo.

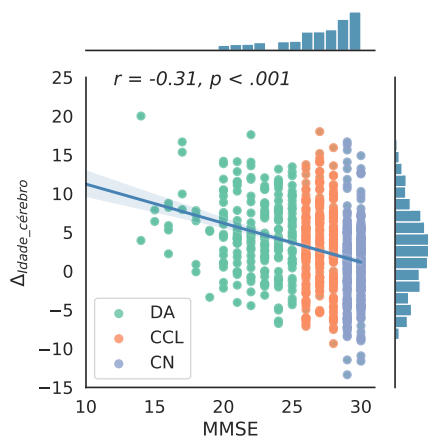


Figura 42 – Gráfico de correlação do $\Delta_{\text{Idade cérebro}}$ e MMSE.

Análise visual qualitativa

De modo a compreender as diferenças entre os grupos de diagnóstico e correlacioná-las com as diferenças de idade, foram criadas imagens médias dos três grupos de diagnóstico e de duas faixas etárias, entre 70 a 77 e entre 78 a 85. A Figura 43 mostra as imagens médias dos três grupos e a imagem de diferença obtida pela subtração entre os pares dos grupos.

Avaliando as imagens médias de cada grupo, é possível observar diferenças sutis, como o aumento da presença do CSF à medida que aumenta a gravidade da DA. Avaliando as imagens de subtração entre os pares, as diferenças se tornam ainda mais visíveis. A CN-DA mostrou diferenças mais significativas nas áreas do hipocampo, evidenciadas pela maior presença da cor vermelha. A imagem CN-CCL também mostrou diferenças, mas em menor grau. Por último, a CCL-DA apresentou diferenças muito mais sutis.

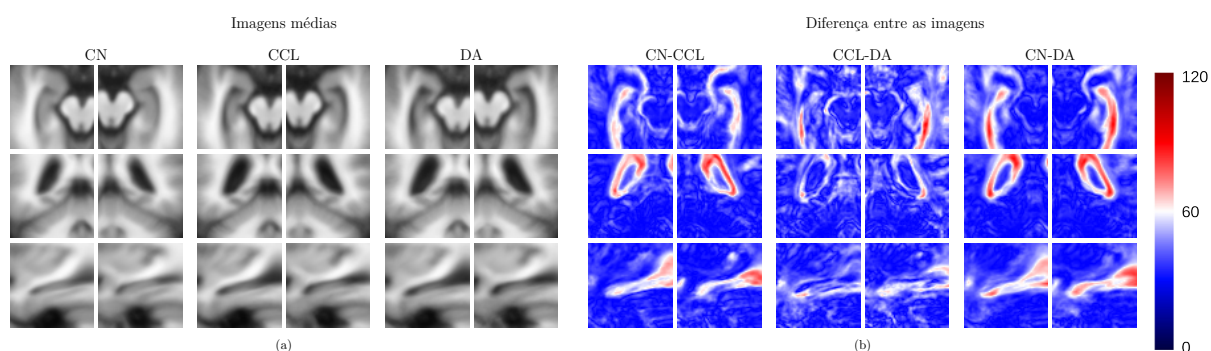


Figura 43 – Imagens de diferença entre pacientes CN e DA, CN e CCL, e CCL e DA para os hipocampos esquerdo e direito nas vistas axial, coronal e sagital, respectivamente.

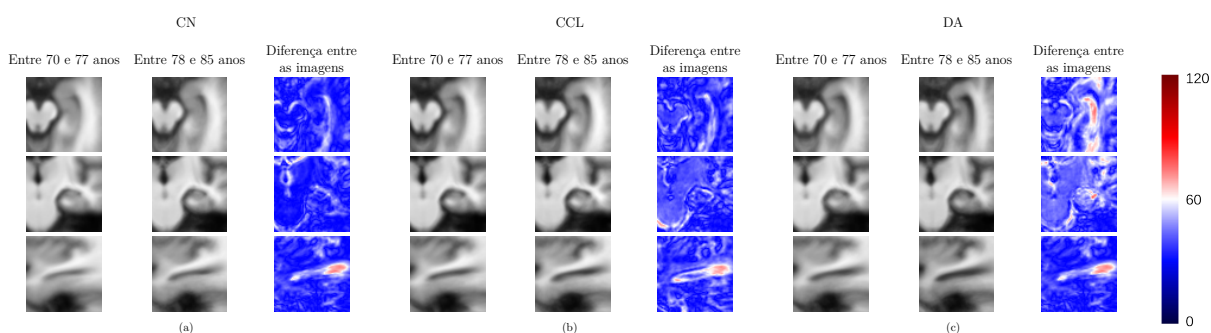


Figura 44 – Imagens de diferença entre as imagens médias de cada grupo de diagnóstico divididas por faixas etárias entre 70 e 77 anos e entre 76 e 85 anos: o grupo CN, CCL e DA para o hipocampo esquerdo.

A Figura 44 mostra as imagens médias dos três grupos de diagnóstico agrupadas por faixa etária e as diferenças entre as faixas por grupo de diagnóstico. A partir dessas imagens, observa-se que o cérebro desenvolve atrofia com o envelhecimento (imagens CN) do paciente e as atrofia são mais acentuadas com a neurodegeneração (imagens CCL e DA). Além disso, as atrofia aumentam em grupos de idade com maior gravidade da doença.

Ao comparar as Figuras 43 e 44 com os *boxplots* apresentados na Figura 41, nota-se que o $\Delta_{\text{Idade cérebro}}$ consegue uma maior diferenciação entre os grupos CN e DA, seguido pelo CN e CCL e, por último, CCL e DA. Essas tendências visuais apresentadas são esperadas e conhecidas na literatura (FRANKE et al., 2010; GASER et al., 2013). Mais importante, elas são consistentes com nossos resultados,

visto que foi obtido um $\Delta_{\text{Idade cérebro}}$ menor para o CN do que para o CCL e menor no CCL quando comparado à DA.

Por fim, em relação ao tempo de treinamento e de inferência dos modelos, para o primeiro conjunto de experimentos, o treinamento do hipocampo direito e esquerdo levou 3 horas e 20 minutos cada, e o tempo de inferência para o *ensemble* levou apenas 0,12 segundos na GPU e 0,165s na CPU. Para o segundo experimento, o treinamento levou 56 minutos para cada hipocampo, e o tempo de inferência do *ensemble* foi o mesmo do primeiro experimento.

Classificação

Para verificar o poder preditivo do $\Delta_{\text{Idade cérebro}}$, o atributo foi inserido em um classificador SVM linear. Assim, como nas abordagens anteriores, foram realizados experimentos pareados entre os grupos de diagnóstico, CN×CCL, CCL×DA e CN×DA. O hiperparâmetro C foi estimado utilizando a metodologia *nested cross-validation* com 10 *folds*, considerando a AUC como critério de desempenho. Além disso, foi utilizado um *grid-search* com o valor 2^x , em que seu expoente x varia entre $[-5, 10]$ com passo de 0,25.

Os resultados de médios de classificação e seus respectivos desvios-padrão para todos os experimentos são apresentados na Tabela 22.

Tabela 22 – Resultados da classificação das imagens utilizando *nested cross-validation* com 10 *folds* (média \pm desvio).

	CN×CCL	CCL×DA	CN×DA
AUC	0,64 \pm 0,07	0,58 \pm 0,08	0,73 \pm 0,09
Acc(%)	58,96 \pm 5,26	56,3 \pm 6,97	65,96 \pm 4,74
Sens(%)	56,52 \pm 9,01	51,83 \pm 12,25	65,95 \pm 11,27
Espec(%)	60,92 \pm 8,27	61,62 \pm 13,56	61,91 \pm 7,49
F1	0,55 \pm 0,06	0,56 \pm 0,08	0,61 \pm 0,07

A partir dos resultados é possível observar o mesmo padrão de diferenciação apresentado pelos *boxplots* (Figura 41) e Figuras 43 e 44. Devido à utilização de somente um atributo para a classificação, os resultados foram inferiores em relação aos obtidos pelas demais abordagens apresentadas. Foram obtidos valores de AUC

de respectivamente 0,64, 0,58 e 0,73 para os experimentos CN×CCL, CCL×DA e CN×DA, acurácia de 58,96, 56,3 e 65,96 e valores de sensibilidade e especificidade bastante similares para todos os experimentos.

5.4 União das abordagens

Conforme descrito no Capítulo 4, as probabilidades das predições que obtiveram o melhor resultado em cada abordagem foram combinadas (*ensemble*) para criar a predição final. Nesse caso, para a abordagem de Atrofia foram utilizados os resultados obtidos da concatenação de todos os atributos (GM|WM|CSF) apresentados nas Tabelas 9, 10 e 11. Já para a abordagem de Assimetria foram utilizados melhores resultados obtidos da concatenação dos atributos Média&Var&Kurt apresentados na Tabela 14 para os experimentos CN×CCL e CN×DA e os resultados da concatenação dos atributos Var&Kurt apresentados na Tabela 13 para o experimento CCL×DA. Por fim, para a abordagem do $\Delta_{\text{Idade c\u00e9rebro}}$ foram utilizados os resultados de classificação apresentados na Tabela 22.

Os resultados para o *ensemble* das abordagens foram obtidos utilizando *nested cross-validation* com 10 *folds* considerando a AUC como crit\u00e9rio de desempenho e os hiperpar\u00e2metros otimizados correspondem aos pesos de cada abordagem para o c\u00e1lculo da predi\u00e7\u00e3o (Equa\u00e7\u00e3o 44). Os pesos ($\text{Peso}_{\text{atrofia}}$, $\text{Peso}_{\text{assimetria}}$ e $\text{Peso}_{\text{delta}}$) foram variados individualmente com valores entre [0, 1] com passos de 0,04. Devido \u00e0 grande quantidade de combina\u00e7\u00f5es dos valores, foram mantidas apenas as que obtiveram somat\u00f3rio unit\u00e1rio.

Para fins comparativos, foram realizados experimentos combinando as abordagens, inicialmente duas a duas e por fim, as tr\u00eas juntas. Os resultados m\u00e9dios de classifica\u00e7\u00e3o e seus respectivos desvios-padr\u00e3o para todos os experimentos (CN×CCL, CCL×DA e CN×DA) s\u00e3o apresentados na Tabela 23. Os melhores resultados (para cada experimento) est\u00e3o indicados em negrito.

A partir dos resultados apresentados \u00e9 poss\u00edvel verificar que, de maneira geral, o uso das predi\u00e7\u00f5es de Atrofia&Assimetria trouxeram maiores ganhos. Analisando a contribui\u00e7\u00e3o de cada abordagem para os resultados finais de classifica\u00e7\u00e3o, observa-se que para a combina\u00e7\u00e3o Atrofia&Assimetria, a Atrofia contribui em mais de 70% para os experimentos CN×CCL e CN×DA em rela\u00e7\u00e3o \u00e0 Assimetria. Por

Tabela 23 – Resultados da classificação das imagens utilizando *nested cross-validation* com 10 *folds* (média \pm desvio) para a união das abordagens.

		Atrofia&Assimetria	Atrofia&Delta	Assimetria&Delta	Atrofia&Assimetria&Delta
CN×CCL	Pesos				
	AUC	0,83 \pm 0,05	0,83 \pm 0,06	0,76 \pm 0,06	0,83 \pm 0,06
	Acc(%)	75,59 \pm 4,84	73,43 \pm 4,78	70,32 \pm 6,21	75,03 \pm 5,31
	Sens(%)	74,34 \pm 10,13	73,25 \pm 11,23	66,09 \pm 13,58	74,92 \pm 8,34
	Espec(%)	75,86 \pm 7,21	73,51 \pm 8,75	73,84 \pm 8,08	75,13 \pm 9,01
	F1	0,74 \pm 0,06	0,71 \pm 0,06	0,66 \pm 0,09	0,73 \pm 0,05
CCL×DA	Pesos				
	AUC	0,76 \pm 0,07	0,73 \pm 0,06	0,69 \pm 0,08	0,75 \pm 0,07
	Acc(%)	69,57 \pm 7,67	67,61 \pm 7,06	65,0 \pm 8,23	68,91 \pm 7,94
	Sens(%)	66,14 \pm 11,8	67,35 \pm 10,86	61,77 \pm 8,78	68,12 \pm 11,93
	Espec(%)	73,62 \pm 8,79	68,02 \pm 9,91	68,86 \pm 10,02	69,76 \pm 10,53
	F1	0,7 \pm 0,09	0,69 \pm 0,08	0,66 \pm 0,08	0,7 \pm 0,09
CN×DA	Pesos				
	AUC	0,96 \pm 0,02	0,95 \pm 0,03	0,9 \pm 0,04	0,96 \pm 0,02
	Acc(%)	89,03 \pm 4,27	88,46 \pm 4,73	80,83 \pm 4,58	89,42 \pm 3,62
	Sens(%)	88,05 \pm 6,81	88,02 \pm 8,2	78,93 \pm 8,21	87,57 \pm 7,15
	Espec(%)	89,73 \pm 6,83	88,72 \pm 6,54	82,16 \pm 7,57	90,73 \pm 6,7
	F1	0,87 \pm 0,05	0,86 \pm 0,06	0,77 \pm 0,05	0,87 \pm 0,04

outro lado, para o experimento CCL×DA, as contribuições foram semelhantes. Para a combinação Atrofia&Delta, a Atrofia contribuiu em mais de 59% em todos os experimentos, e quase 100% para o experimento CN×DA. Analisando a combinação Assimetria&Delta, as contribuições foram semelhantes para o experimento CN×CCL, enquanto que a Assimetria teve maior contribuição para os demais. Finalmente, para a combinação Atrofia&Assimetria&Delta, observa-se que a Atrofia contribuiu em 49,6%, 55,2% e 39,6%, respectivamente, para os experimentos CN×CCL, CCL×DA e CN×DA, enquanto, para os mesmos experimentos, a Assimetria contribuiu em 20%, 30% e 46,4% e o Delta em 30,4%, 14,8% e 14%.

É importante mencionar que o uso das abordagens individuais foram consideradas no espaço de parâmetros de busca e não foram selecionadas, indicando que

o uso das combinações entre as abordagens contribuíram positivamente em cada experimento, mesmo que seja com baixa ponderação. Além disso, outro ponto a ser mencionado é que os resultados de classificação exibidos neste experimento foram calculados usando a limiarização correspondente ao ponto de corte da curva ROC cuja distância até o ponto $(0, 1)$ do gráfico foi mínima. Esse ponto corresponde ao melhor compromisso entre as métricas de sensibilidade e especificidade, ou seja, a menor diferença entre elas. Portanto, é esperado que haja algumas perdas em relação às métricas de acurácia e F1-score em uma comparação direta com as avaliações das abordagens individuais.

Para realizar a comparação de todos os resultados obtidos foram plotados gráficos de barras, exibidos nas Figuras 45, 46 e 47. Para auxiliar na interpretação, as abordagens foram ordenadas em relação à maior AUC, seguido pela maior acurácia, em caso de empate.

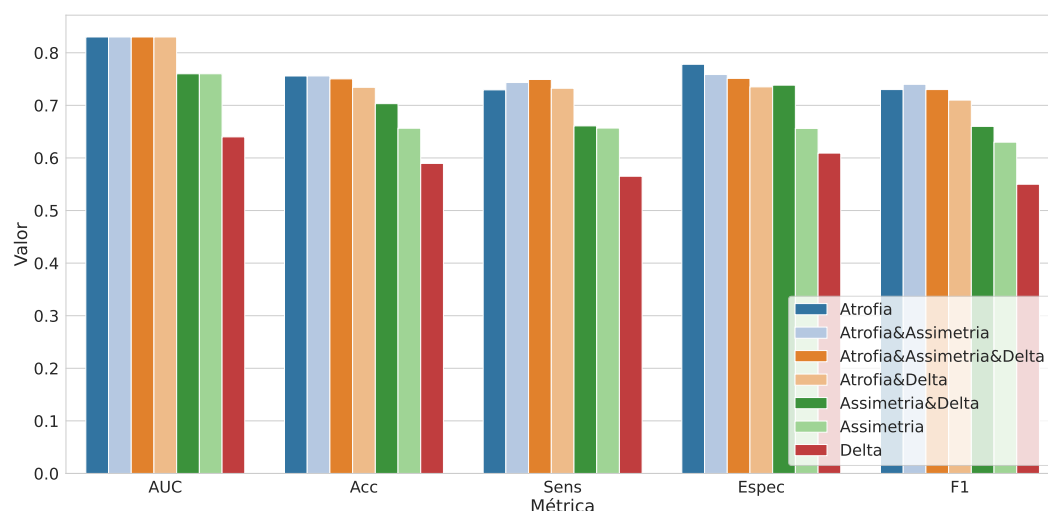


Figura 45 – Gráfico de barras para a comparação dos resultados para o experimento $CN \times CCL$

Para o experimento $CN \times CCL$ é possível observar que todas as abordagens que utilizaram a Atrofia obtiveram resultados superiores. Nesse caso, considerando que as AUCs foram as mesmas, a abordagem de Atrofia (individual) se mostrou superior pelo fato de não depender de outras abordagens para melhorar os resultados. No entanto, um ponto a ser mencionado é a superioridade de 1% no valor do F1-score para a Atrofia&Assimetria e melhor balanceamento entre as métri-

cas de sensibilidade e especificidade em relação à abordagem individual. Além disso, o uso das três abordagens obteve a menor diferença entre a sensibilidade e especificidade, e o menor desvio-padrão entre os *folds*.

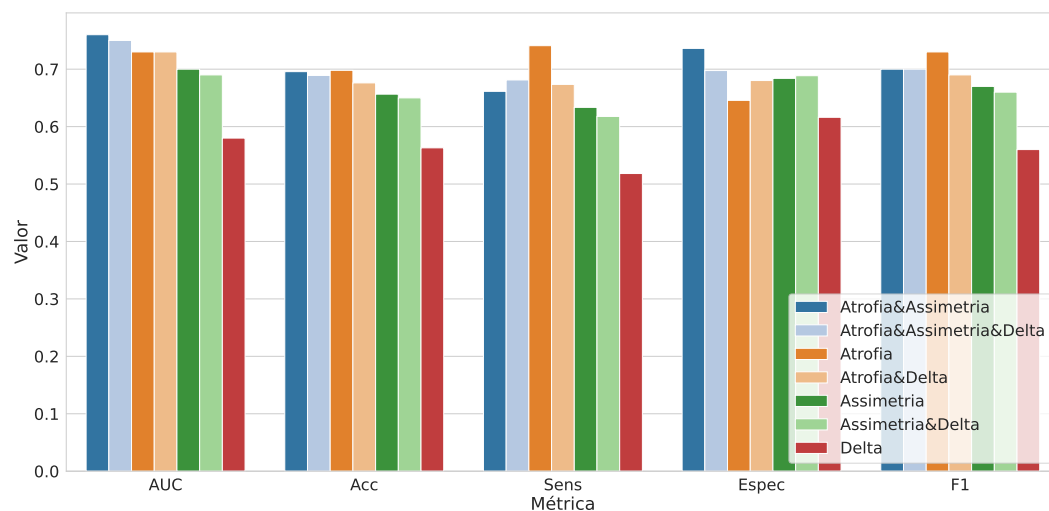


Figura 46 – Gráfico de barras para a comparação dos resultados para o experimento CCL×DA

Para o experimento CCL×DA, também é possível observar que todas as abordagens que utilizaram a Atrofia obtiveram resultados superiores. Nesse caso, a combinação das abordagens Atrofia&Assimetria trouxe maiores ganhos, com superioridade de 3% no valor da AUC em relação à abordagem individual de Atrofia, seguida pela combinação das três abordagens, Atrofia&Assimetria&Delta, que obteve valores um ganho de 2% em relação à Atrofia. Outro ponto a ser considerado é o melhor balanceamento entre as métricas de sensibilidade e especificidade para a combinação das três abordagens.

Para o experimento CN×DA, mais uma vez, é possível observar que todas as abordagens que utilizaram a Atrofia obtiveram resultados superiores. Nesse caso, a combinação de todas as abordagens Atrofia&Assimetria&Delta trouxe maiores ganhos, com superioridade de 1% no valor da AUC e 0,18% na acurácia em relação à abordagem individual de Atrofia, seguida pela combinação das abordagens Atrofia&Assimetria, que obteve um ganho de 1% de AUC em relação à Atrofia. Além disso, o uso do *ensemble* da Atrofia&Assimetria resultou na menor diferença entre a sensibilidade e especificidade, e o menor desvio-padrão entre os *folds*.

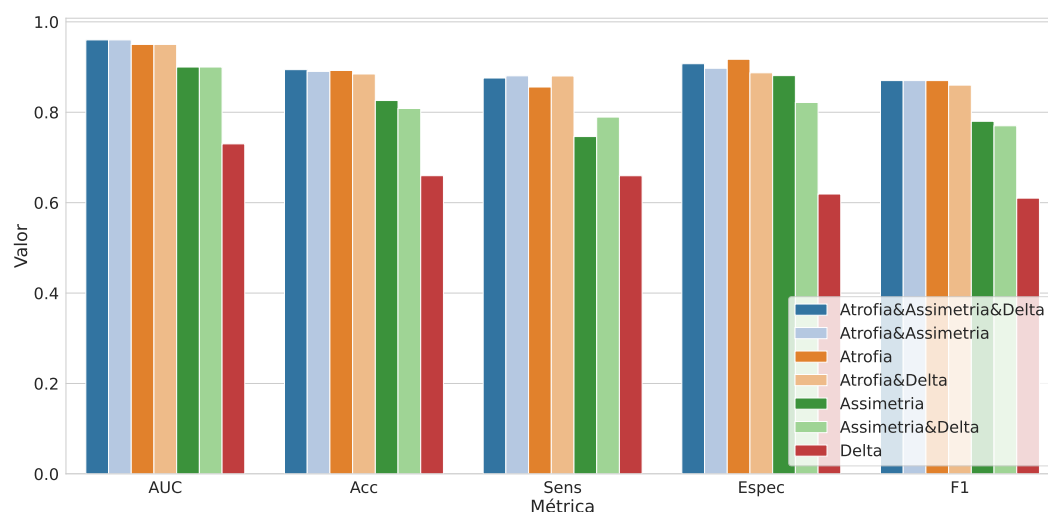


Figura 47 – Gráfico de barras para a comparação dos resultados para o experimento CN×DA

A proposta de unir as abordagens foi considerada assumindo a hipótese de que as abordagens seriam capazes de produzir respostas complementares. No entanto, como o resultado de classificação para a predição da idade do cérebro foi bastante aquém dos demais, era esperado que o uso de suas probabilidades de predição trouxessem pouco ou nenhum ganho. Além disso, visto que a abordagem de atrofia produziu resultados bastante superiores à de assimetrias, também era esperado que suas probabilidades trouxessem contribuições desbalanceadas. De maneira geral, com ponderações distintas, os resultados obtidos com a combinação dos atributos Atrofia&Assimetria e os três, Atrofia&Assimetria&Delta, trouxeram contribuições positivas para os resultados finais de classificação, embora, em alguns casos, o custo é elevado.

5.5 Comparação dos resultados

Os resultados obtidos por este trabalho foram comparados diretamente, quando possível, e indiretamente com os métodos da literatura.

5.5.1 Comparação direta

Para comparar diretamente os resultados obtidos neste trabalho com o de outro método da literatura, visto que há uma grande variação na quantidade de imagens utilizadas nos trabalhos, o método *ROI-based* (subseção 4.8) foi implementado. Para uma comparação justa, assim como nos experimentos apresentados acima, foi utilizado o algoritmo SVM. Foram testados os *kernels* RBF, polinomial e linear com validação cruzada estratificada em 10 *folds*, considerando a AUC como critério de desempenho. Os melhores resultados foram obtidos para o classificador SVM com *kernel* RBF e o melhor valor para o parâmetro C de cada SVM foi determinado via *coarse-to-fine grid-search* com as mesmas variações descritas na seção 5.1.1.

A Tabela 24 apresenta os valores médios de classificação correspondentes aos melhores resultados obtidos das abordagens desenvolvidas neste projeto (Atrofia, Assimetria, Delta e União das abordagens) e os resultados do método *ROI-based*.

Tabela 24 – Resultados da classificação das imagens utilizando *nested cross-validation* com 10 *folds* (média \pm desvio) para cada abordagem.

		Atrofia	Assimetria	Delta	União das abordagens	<i>ROI-based</i>
CN×CCL	AUC	0,83 \pm 0,05	0,76 \pm 0,05	0,64 \pm 0,07	0,83 \pm 0,05	0,8 \pm 0,04
	Acc(%)	75,58 \pm 3,6	65,65 \pm 4,71	58,96 \pm 5,26	75,59 \pm 4,84	72,68 \pm 3,38
	Sens(%)	72,94 \pm 8,41	65,68 \pm 8,41	56,52 \pm 9,01	74,34 \pm 10,13	74,85 \pm 8,84
	Espec(%)	77,81 \pm 6,84	65,61 \pm 7,35	60,92 \pm 8,27	75,86 \pm 7,21	70,78 \pm 10,39
	F1	0,73 \pm 0,04	0,63 \pm 0,06	0,55 \pm 0,06	0,74 \pm 0,06	0,71 \pm 0,039
CCL×DA	AUC	0,73 \pm 0,07	0,7 \pm 0,09	0,58 \pm 0,08	0,76 \pm 0,07	0,74 \pm 0,06
	Acc(%)	69,8 \pm 6,73	65,65 \pm 7,79	56,3 \pm 6,97	69,57 \pm 7,67	67,39 \pm 6,22
	Sens(%)	74,09 \pm 6,3	63,35 \pm 9,9	51,83 \pm 12,25	66,14 \pm 11,8	70,55 \pm 12,43
	Espec(%)	64,57 \pm 11,17	68,4 \pm 8,81	61,62 \pm 13,56	73,62 \pm 8,79	63,57 \pm 7,53
	F1	0,73 \pm 0,06	0,67 \pm 0,08	0,56 \pm 0,08	0,7 \pm 0,09	0,7 \pm 0,06
CN×DA	AUC	0,95 \pm 0,03	0,9 \pm 0,04	0,73 \pm 0,09	0,96 \pm 0,02	0,9 \pm 0,04
	Acc(%)	89,24 \pm 4,04	82,59 \pm 6,84	65,96 \pm 4,74	89,42 \pm 3,62	82,59 \pm 5,19
	Sens(%)	85,58 \pm 8,5	74,62 \pm 9,6	65,95 \pm 11,27	87,57 \pm 7,15	79,93 \pm 6,61
	Espec(%)	91,72 \pm 5,0	88,1 \pm 7,67	61,91 \pm 7,49	90,73 \pm 6,7	84,45 \pm 6,79
	F1	0,87 \pm 0,05	0,78 \pm 0,09	0,61 \pm 0,07	0,87 \pm 0,04	0,79 \pm 0,06

Analisando os resultados de classificação dos dois métodos nota-se que os resultados obtidos pela Atrofia foram superiores aos da técnica *ROI-based* para todos os experimentos, considerando as métricas de acurácia e F1. Em relação à AUC, os resultados superiores foram obtidos para os experimentos de classificação CN×CCL e CN×DA e para o experimento CCL×DA, houve apenas uma pequena diferença entre os valores. Uma possível justificativa pode ser dada pelo fato do experimento CCL×DA ter sido realizado com um atlas de pacientes CN, criando

assim a necessidade de afrouxamento dos parâmetros inicialmente definidos, o que pode ter levado a seleção incorreta de alguns pontos salientes e afetado as demais etapas. No geral, o desempenho de classificação alcançado foi superior, provavelmente causado pelos dois níveis de classificação utilizados e seleção de pontos salientes discriminativos. Além disso, esses resultados foram alcançados apenas para as regiões dos hipocampos, enquanto o ROI-*based* utilizou noventa regiões do cérebro.

Para a abordagem de assimetrias foram obtidos resultados de classificação comparáveis ao do método ROI-*based* para o experimento CN×DA, no entanto, os resultados para os experimentos CN×CCL e CCL×DA foram inferiores. Nesse caso, a justificativa pode ser dada devido à abordagem de assimetrias usar apenas informações assimétricas extraídas das regiões dos hipocampos, as quais, conforme os demais experimentos apresentados, se mostraram muito sutis para a diferenciação da classe CCL.

Para o Delta, os resultados obtidos foram inferiores em todos os experimentos, provavelmente devido ao experimento ter enfoque no erro de estimação da idade do cérebro e usá-lo exclusivamente como atributo. Para a união das abordagens, os resultados foram superiores em todos os experimentos e métricas — devido à utilização das probabilidades de predição que se mostraram complementares.

A Figura 48 mostra as curvas ROC e seus valores de AUC correspondentes para a classificação CN×CCL, CCL×DA e CN×DA. Nesse caso, as curvas ROC para o Delta foram omitidas pelo fato de apresentar resultados inferiores. A partir da análise das curvas nota-se que apesar de existir pouca diferença entre as curvas para o experimento CN×CCL é possível verificar o melhor desempenho para a abordagem de Atrófias. Nas curvas dos experimentos CN×DA e CCL×DA, a União das abordagens apresentou maior destaque.

Considerando que o método ROI-*based* é frequentemente usado como “método controle” em artigos publicados na literatura, pode-se concluir que a abordagem de Atrófias e a proposta de União das abordagens produziram resultados consideravelmente maiores que o ROI-*based*, indicando que a avaliação volumétrica de estruturas cerebrais pode não ser suficiente para a detecção das alterações estruturais geradas no desenvolvimento da DA.

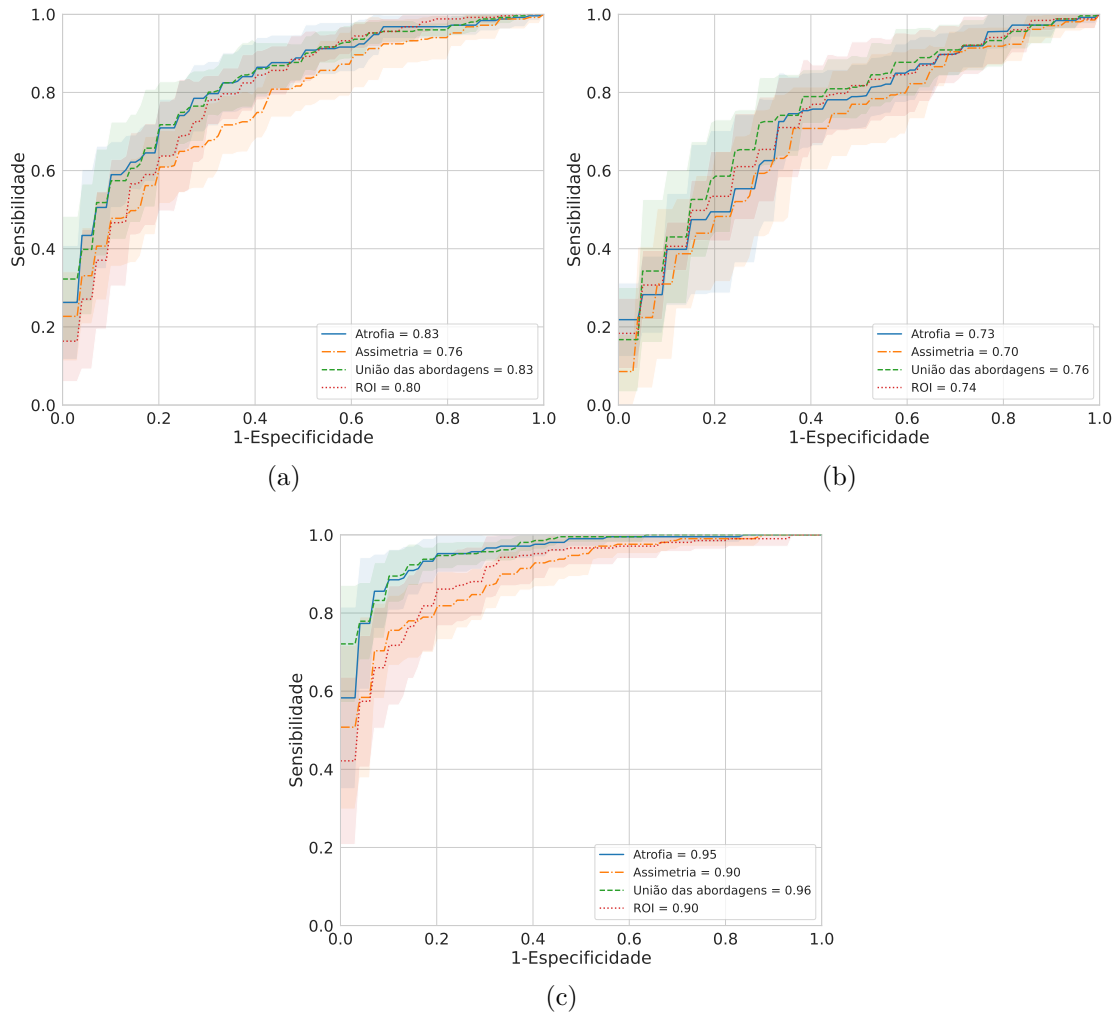


Figura 48 – Curvas ROC e valores de AUC correspondentes aos resultados de classificação de (a) $CN \times CCL$, (b) $CCL \times DA$ e (c) $CN \times DA$ para os métodos de atrofia, assimetria, união das abordagens e ROI-based.

5.5.2 Comparação indireta

Para uma comparação não direta, a Tabela 25 é fornecida. Nela, são apresentados resultados do estado da arte em classificação na DA usando imagens de RM estruturais.

A partir da Tabela de comparação, é possível observar que há uma diferença considerável na quantidade de imagens utilizadas pelos trabalhos e que os resul-

Tabela 25 – Comparação indireta de nossos resultados sobre a classificação de imagens de RM de casos CN×CCL, CCL×CN e CN×AD com estudos relacionados.

Trabalho	# Pacientes			CN×CCL				CCL×DA				CN×DA			
	CN	MCI	AD	Acc(%)	Sens(%)	Spec(%)	AUC	Acc(%)	Sens(%)	Spec(%)	AUC	Acc(%)	Sens(%)	Spec(%)	AUC
Liu et al. (2014)	77	169	65	76,92	74,29	78,13	-	-	-	-	-	87,76	88,57	87,22	-
Ye et al. (2016)	52	99	51	68,02	76,87	51,15	0,71	-	-	-	-	87,26	88,43	86,15	0,93
Zhang et al. (2016)	430	731	358	73,6	75,3	69,7	-	-	-	-	-	83,7	80,5	85,1	-
Korolev et al. (2017)	61	43 (pMCI)	50	63,00	-	-	0,67	62,00	-	-	0,61	79,0	-	-	0,88
Zhang et al. (2017)	207	346	154	79,02	90,46	59,9	0,85	-	-	-	-	88,3	79,61	94,69	0,94
Cao et al. (2017)	229	397	192	71,9	79,0	66,7	0,73	74,7	69,1	79,1	0,76	88,6	85,7	90,4	0,9
Liu et al. (2018)	429	-	358	-	-	-	-	-	-	-	-	91,09	88,05	95,3	0,96
Lian et al. (2020)	429	-	358	-	-	-	-	-	-	-	-	90,3	82,4	96,5	0,95
Liu et al. (2020)	119	233	97	76,2	79,5	69,8	0,775	-	-	-	-	88,9	86,6	90,8	0,92
Abordagem de atrofas	302	251	209	75,58	72,94	77,81	0,83	69,8	74,09	64,57	0,73	89,24	85,58	91,72	0,95
União das abordagens	302	251	209	75,59	74,34	75,86	0,83	69,57	66,14	73,62	0,76	89,42	87,57	90,73	0,96

tados obtidos nesta pesquisa estão muito próximos aos obtidos pelo estado da arte.

5.6 Considerações finais

Este capítulo apresentou e discutiu os resultados obtidos nas abordagens desenvolvidas nesta pesquisa, todas objetivaram a classificação das imagens entre as classes CN×CCL, CCL×DA e CN×DA. Na análise de atrofas estruturais foram apresentados e discutidos os resultados para a classificação no nível de pontos salientes e das imagens. Além disso, foram apresentados os resultados obtidos em cada etapa, as configurações experimentais e a avaliação dos parâmetros utilizados. A partir dos resultados, notou-se que a proposta de concatenação dos atributos influenciou positivamente os resultados dos três experimentos. Foram realizadas também análises qualitativas para auxiliar no entendimento e interpretação dos resultados da abordagem. Por fim, na comparação direta com o ROI-*base*, nosso método obteve desempenho de classificação superior para a maioria das métricas avaliadas.

Na análise de assimetrias estruturais foram apresentados e discutidos os resultados da classificação das imagens e a análise estatística do índice *IA*, criado para verificar se cada atributo estatístico proposto possuía distinção estatisticamente significativa entre os grupos de diagnóstico. Foram apresentadas também as configurações experimentais e a avaliação dos parâmetros utilizados. Os resultados dos experimentos foram bastante promissores, principalmente por usarem apenas atri-

butos assimétricos extraídos dos hipocampos. Notou-se que a concatenação dos atributos influenciou positivamente nos resultados e as concatenações contendo o atributo de Variância proporcionaram os maiores ganhos. Além disso, o *IA* apresentou a mesma tendência de assimetria relatada em trabalhos de neurologia, com valores de assimetria aumentando a medida em que a condição neurodegenerativa se agrava.

Na análise do erro de predição da idade biológica do cérebro foram apresentadas as configurações experimentais e os resultados estatísticos, de regressão e de classificação, além de análises qualitativas. Os resultados mostraram que o uso do aumento dos dados foi muito importante para a convergência da rede. Além disso, o uso de duas etapas de treinamento forneceu análises independentes e complementares, que auxiliaram no bom desempenho do treinamento de ambos os experimentos. Os resultados de regressão obtidos são competitivos em relação a outros métodos relatados na literatura (Tabela 4), mesmo usando somente a região do hipocampo. Ademais, as conclusões estatísticas mostraram uma tendência para o $\Delta_{\text{Idade cérebro}}$ e as outras métricas consistentes com o avanço do declínio cognitivo: os valores são mais baixos no CN do que no CCL e no CCL do que na DA sendo encontrada uma correlação significativa e negativa entre o MMSE e o $\Delta_{\text{Idade cérebro}}$. Por outro lado, os resultados de classificação foram aquém dos relatados na literatura.

Por fim, na proposta de unir as abordagens foram apresentadas as configurações experimentais e os resultados. Nesse caso, notou-se que, a depender do experimento, as respostas obtidas das abordagens foram complementares e auxiliaram no aumento das métricas de avaliação. De forma geral, a abordagem de *Atrofias* apresentou uma contribuição mais significativa que as demais. As conclusões de todas as abordagens são apresentadas no Capítulo 6

Capítulo 6

Conclusões

Este capítulo sintetiza as principais conclusões da revisão da literatura e os principais resultados obtidos pelas abordagens desenvolvidas. Além disso, também indica novas investigações que podem ser derivadas deste estudo.

6.1 Análise de atrofas estruturais

Esta abordagem criou uma nova metodologia para realizar a classificação das imagens de RM para auxiliar no diagnóstico da DA usando pontos salientes discriminativos entre populações distintas, CN, CCL e DA. Especificamente, foi projetada uma estrutura de classificação em dois níveis, baseada em pontos salientes e em informações gerais da imagem. O objetivo da abordagem em dois níveis foi de aumentar o número de amostras de treinamento no primeiro nível e criar uma “margem de segurança” que permitisse bom desempenho no segundo nível, mesmo com a ocorrência de erros de classificação no nível de pontos.

Em geral, os métodos publicados na literatura buscam identificar mudanças estruturais nas imagens do cérebro avaliando *voxels*, ROIs e *patches*. Muitos deles extraem uma quantidade elevada de atributos e criam espaços de atributos de alta dimensionalidade, gerando um aumento na complexidade dos classificadores

e diminuindo a sua interpretabilidade. Além disso, devido ao número limitado de imagens normalmente disponíveis com doenças cognitivas, questiona-se se o fenômeno de *overfitting* não estaria presente no treinamento dos classificadores. Diferentemente dos trabalhos analisados, uma preocupação desta pesquisa foi a de manter uma baixa dimensionalidade do espaço de atributos, visto o reduzido número de imagens. Desta maneira, nosso método focou primeiro na seleção automática de pontos salientes mais discriminantes entre populações que representassem mudanças cerebrais causadas por patologias do CCL ou da DA. Para isso, foi proposto o uso de um atlas de pontos salientes, obtidos de uma população CN, como referência. O uso do atlas, além de reduzir o número de pontos avaliados, também ajuda a eliminar pontos instáveis ou ruidosos, diminuindo, portanto, o número de comparações. Em seguida, foram utilizados os pontos salientes dos três grupos em experimentos pareados para treinar os classificadores desse nível. Por fim, foi criado um pequeno vetor de atributos, que é interpretável, para realizar a classificação das imagens no segundo nível, a partir das predições dos pontos salientes.

Foram obtidos resultados bastante promissores e consistentes em todas as etapas de classificação com padrões condizentes com a literatura médica. O método foi diretamente comparado com o ROI-*based* e um obteve desempenho de classificação superior para a maioria das métricas avaliadas. Além disso, foram obtidos resultados similares a outros estudos (Tabela 2) presentes na literatura. Devido à utilização de bases de dados e quantidade de imagens distintas, os resultados podem apenas ser discutidos com observações indiretas: *i*) o número de imagens utilizados para avaliar esta abordagem foi superior à alguns dos trabalhos; *ii*) esta abordagem utilizou validação cruzada aninhada, garantindo partições independentes para os conjuntos de treinamento, validação e teste, enquanto muitos métodos publicados não fazem uso de um conjunto de validação; *iii*) os resultados foram obtidos usando apenas as regiões dos hipocampos, enquanto a maioria dos outros métodos usou todo o cérebro; *iv*) e os resultados obtidos superaram a maioria dos métodos e são muito semelhantes aos outros.

Como trabalhos futuros, novos estudos devem ser conduzidos para aumentar os resultados de classificação para os casos CN×CCL e CCL×DA e aumentar a classificação no nível de pontos salientes, o que conseqüentemente, também aumentaria

os resultados no nível da imagem. Além disso, devido à modelagem da abordagem possuir experimentos independentes para a maximização dos parâmetros, novos descritores e novas imagens como fontes de informação podem ser exploradas, além de novas ROIs. Ademais, visto que este foi um trabalho multidisciplinar, pretendemos envolver especialistas médicos na avaliação dos resultados.

6.2 Análise de assimetrias estruturais

Nesta abordagem, foi criada uma nova técnica para realizar a classificação de imagens de RM, também para o auxílio ao diagnóstico DA, usando apenas atributos hipocampais assimétricos provindos de respostas direcionais via convolução com filtros log-Gabor 3D. Além disso, foi desenvolvido um novo índice de assimetria que apresentou conclusões estatísticas condizentes com a literatura médica.

A maioria dos métodos propostos na literatura que exploram assimetrias hipocampais em imagens de RM no contexto da DA (SARICA et al., 2018; ARDEKANI et al., 2019), em geral, utilizam diferenças entre o volume dos hipocampus esquerdo e direito como principal atributo, o que os tornam dependentes do sucesso da técnica utilizada para a segmentação das imagens. Além disso, esses trabalhos se concentram em desenvolver biomarcadores para a progressão da DA, portanto, não verificam o comportamento desses atributos assimétricos para a classificação. Por outro lado, esta abordagem extrai atributos assimétricos por meio da filtragem de imagens e os utiliza para a classificação das imagens de RM.

Embora os resultados obtidos não sejam tão altos quanto os relatados em outros estudos (Tabela 2), é importante enfatizar que são utilizados apenas atributos que representam assimetrias entre os hipocampus para a classificação das imagens. Além disso, conforme mencionado na abordagem anterior, *i*) o número de imagens utilizados foi superior à alguns dos trabalhos; *ii*) esta abordagem utilizou validação cruzada aninhada; *iii*) e os resultados foram obtidos usando apenas para as regiões dos hipocampus. Outro ponto importante a salientar é que as conclusões estatísticas obtidas para o índice de assimetria são condizentes com a literatura médica, ou seja, o método encontrou assimetria maior em homens do que em mulheres (ARDEKANI et al., 2019; LUCARELLI et al., 2013) e aumento

da assimetria inter-hemisférica hipocampal com o aumento do declínio cognitivo (SARICA et al., 2018; ARDEKANI et al., 2019; SHI et al., 2009).

Como trabalhos futuros, novas investigações devem ser conduzidas para explorar novos atributos para caracterizar as assimetrias hipocampais. Além disso, precisam ser estudadas técnicas e métodos para verificar e encontrar atributos com maior capacidade de distinção entre os grupos CCL e DA. Por fim, mais uma vez, podem ser analisadas diferentes ROIs, visto que as imagens foram filtradas para todo o encéfalo, permitindo que outras regiões sejam facilmente analisadas.

6.3 Análise do erro de predição da idade biológica do cérebro

Nesta abordagem foi criado um método para estimar a idade biológica do cérebro (regressão) de pacientes CN usando uma arquitetura de rede neural convolucional 3D eficiente, baseada na arquitetura desenvolvida para funcionar com dispositivos móveis (SANDLER et al., 2018). O método usa o erro da estimação para a realização de análises estatísticas em diferentes grupos de diagnóstico e como atributo em experimentos de classificação. Foram utilizadas algumas técnicas para o aumento dos dados com o intuito de obter uma distribuição mais balanceada das faixas etárias e melhorar os resultados da regressão.

Foram também realizados treinamentos em duas etapas que resultaram em duas análises individuais. O tempo de inferência dos dois modelos foi de 0,12s e o tempo total de processamento de uma imagem foi de menos de sete minutos. No primeiro treinamento foram alcançadas métricas bastante competitivas para a regressão (Tabela 4 fornecida para comparação) sendo fornecida uma análise qualitativa dos resultados. A segunda etapa forneceu uma análise da significância estatística do $\Delta_{\text{Idade cérebro}}$ usando três grupos distintos (CN, CCL e DA) com efeitos de envelhecimento distintos e estágios de doenças neurológicas. Além disso, os resultados foram corroborados com medições clínicas, como o MMSE, indicando a possibilidade de usar o $\Delta_{\text{Idade cérebro}}$ como um biomarcador para identificar e auxiliar no diagnóstico do CCL e da DA. Por fim, o $\Delta_{\text{Idade cérebro}}$ foi utilizado como atributo unitário para a classificação e obteve resultados consistentes com os

demais, embora foram aquém dos apresentados na literatura para a classificação.

Como trabalhos futuros, sugere-se uma análise por ROI, visto que os hipocampus já obtiveram resultados tão promissores. Além de melhorar os resultados da regressão, novas ROIs tendem a melhorar os resultados de classificação, com a inserção de mais atributos descritivos. Além disso, devido ao aspecto longitudinal do conjunto de dados ADNI, novos estudos devem ser conduzidos para determinar se o $\Delta_{\text{Idade c\u00e9rebro}}$ poderia capturar as diferen\u00e7as apresentadas pelas imagens ao longo do tempo (geralmente repetidas a cada seis meses).

6.4 Uni\u00e3o das abordagens

Nesta etapa, as probabilidades das predi\u00e7\u00f5es das etapas anteriores foram combinadas com o objetivo de melhorar os resultados individuais obtidos em cada uma delas. Visto que os resultados obtidos fazem uso das probabilidades individuais e o padr\u00e3o de valor das probabilidades segue a ordem *Atrofia > Assimetria >> Delta*, ent\u00e3o \u00e9 esperado que as contribui\u00e7\u00f5es sejam desbalanceadas entre as abordagens. Os resultados obtidos foram comparados com as abordagens individuais e o m\u00e9todo implementado para a compara\u00e7\u00e3o direta dos resultados, o *ROI-based*.

Considerando a compara\u00e7\u00e3o dos melhores resultados da uni\u00e3o das abordagens (em rela\u00e7\u00e3o \u00e0 AUC — $\text{CN} \times \text{CCL}$ e $\text{CCL} \times \text{DA}$: *Atrofia&Assimetria*, $\text{CN} \times \text{DA}$: *Atrofia&Assimetria&Delta*) com o *ROI-based*, a uni\u00e3o obteve desempenho de classifica\u00e7\u00e3o superior \u00e0s abordagens individuais para todas as m\u00e9tricas avaliadas. Al\u00e9m disso, em uma compara\u00e7\u00e3o indireta com a literatura (Tabela 2), obteve resultados superiores em alguns casos e bastante competitivos em outros.

Ao realizar uma segunda an\u00e1lise dos resultados dos experimentos, foram considerados o custo do uso de uma abordagem a mais e dois fatores principais, o valor da AUC e a curva ROC que melhor divide os dados sens\u00edveis e espec\u00edficos, visto que as demais m\u00e9tricas podem ser ajustadas e outros valores obtidos. Para a classifica\u00e7\u00e3o $\text{CN} \times \text{CCL}$, a abordagem de *Atrofias* (individual) foi a mais indicada, visto que obteve o mesmo valor de AUC que as combina\u00e7\u00f5es e \u00e9 independente das outras abordagens. Para a classifica\u00e7\u00e3o $\text{CCL} \times \text{DA}$, a abordagem de *Atrofia&Assimetria* trouxe um ganho consider\u00e1vel no valor da AUC, evidenciando que as duas abordagens produziram respostas complementares. Por fim, para o experimento $\text{CN} \times \text{DA}$,

embora a combinação das três abordagens, Atrofia&Assimetria&Delta, tenha resultado numa melhor acurácia, a combinação Atrofia&Assimetria obteve o mesmo valor de AUC e a curva ROC que melhor divide os dados sensíveis e específicos com o menor desvio-padrão, além de utilizar uma abordagem a menos.

6.5 Considerações finais

Este capítulo apresentou as conclusões obtidas individualmente para cada abordagem proposta e desenvolvida nesta Tese, assim como uma comparação com a literatura. Todos os aspectos de originalidade e questões de pesquisa levantados no Capítulo 1 foram alcançados, além de todas as abordagens terem sido publicadas em periódicos de alto impacto. As possibilidades de melhorias a serem exploradas foram mencionadas nas sugestões de trabalhos futuros. Sucintamente, conseguimos produzir três metodologias distintas que se complementam, se necessário. Devido à somente as abordagens individuais de atrofia e assimetria possuírem enfoques em classificação, a combinação das mesmas produziu os melhores resultados. Por fim, conforme mostrado, os resultados de classificação obtidos por este trabalho se posicionam entre os melhores da literatura e podem contribuir positivamente para a evolução de novas técnicas e estudos relacionados ao desenvolvimento de DA e auxílio ao diagnóstico da doença.

Referências

- ABBOTT, A. Dementia: A problem for our age. **Nature**, v. 475, n. 7355, p. S2–S4, 2011.
- ABDI, H.; WILLIAMS, L. J. Tukey’s Honestly Significant Difference (HSD) test. **Encyclopedia of Research Design**, v. 3, n. 1, p. 1–5, 2010.
- ABE, S. **Support vector machines for pattern classification**. 2. ed. Secaucus, NJ, USA: Springer, 2005. 21-112 p.
- AHMED, O. B. et al. Classification of Alzheimer’s disease subjects from MRI using hippocampal visual features. **Multimedia Tools and Applications**, v. 74, n. 4, p. 1249–1266, 2015.
- AHMED, O. B. et al. Alzheimer’s disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. **Computerized Medical Imaging and Graphics**, v. 44, p. 13–25, 2015.
- ALZHEIMER’S DISEASE INTERNATIONAL. **The global voice on Dementia**. Londres, Reino Unido, 2020. Disponível em: <<https://www.alz.co.uk/>>.
- ANKERST, M. et al. 3D shape histograms for similarity search and classification in spatial databases. In: **International Symposium on Spatial Databases**. Hong Kong, China: Springer, 1999. p. 207–226.
- ARAÚJO, M. R. M.; POLONI, K. M.; FERRARI, R. J. Assessment of linear and non-linear feature projections for the classification of 3-D MR images on Cognitively Normal, Mild Cognitive Impairment and Alzheimer’s disease.

In: **International Conference on Computational Science and Its Applications**. Cham: Springer International Publishing, 2021. p. 131–146.

ARDEKANI, B. A. et al. Sexual dimorphism and hemispheric asymmetry of hippocampal volumetric integrity in normal aging and Alzheimer disease. **American Journal of Neuroradiology**, v. 40, n. 2, p. 276–282, 2019.

ASHBURNER, J. et al. Identifying global anatomical differences: Deformation-based morphometry. **Human Brain Mapping**, v. 6, n. 5-6, p. 348–357, 1998.

ASSOCIATION, A. Alzheimer's disease facts and figures. **Alzheimer's & dementia: The Journal of the Alzheimer's Association**, v. 16, n. 3, p. 391, 2020.

BARNES, J. et al. Does Alzheimer's disease affect hippocampal asymmetry? Evidence from a cross-sectional and longitudinal volumetric MRI study. **Dementia and Geriatric Cognitive Disorders**, v. 19, n. 5-6, p. 338–344, 2005.

BASSIONY, H. S. et al. Comparative histological study on the therapeutic effect of green tea and stem cells in Alzheimer's disease complicating experimentally induced diabetes. **International Journal of Stem Cells**, v. 8, n. 2, p. 181–190, 2015.

BAUMGARTNER, C. F. et al. Phiseg: Capturing uncertainty in medical image segmentation. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI**. Shenzhen, China, 2019. p. 119–127.

BEACH, T. G. et al. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer disease centers, 2005–2010. **Journal of neuropathology and experimental neurology**, v. 71, n. 4, p. 266–273, 2012.

BELL, S. et al. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition - CVPR**. Las Vegas, USA: IEEE, 2016. p. 2874–2883.

BELONGIE, S.; MALIK, J.; PUZICHA, J. Shape matching and object recognition using shape contexts. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 4, p. 509–522, 2002.

- BENGIO, Y. Practical recommendations for gradient-based training of deep architectures. In: **Neural networks: Tricks of the trade**. Berlin, Heidelberg: IEEE, 2012. p. 437–478.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. 1. ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc, 2006. 326-344 p.
- BLOCH, F. Nuclear induction. **Physical review**, v. 70, n. 7-8, p. 460, 1946.
- BOULANGER, J.; KERVRANN, C.; BOUTHEMY, P. Adaptive spatio-temporal restoration for 4D fluorescence microscopic imaging. In: **International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI**. Cham: Springer International Publishing, 2005. p. 893–901.
- BUADES, A.; COLL, B.; MOREL, J.-M. A review of image denoising algorithms, with a new one. **Multiscale Modeling & Simulation**, v. 4, n. 2, p. 490–530, 2005.
- CAMBUI, V. H. C.; POLONI, K. M.; FERRARI, R. J. Classification of brain MR images for the diagnosis of Alzheimer’s disease based on features extracted from the three main brain tissues. In: **Proceedings of the 6th Brazilian Technology Symposium (BTSym’20)**. Cham: Springer International Publishing, 2021. p. 212–219.
- CAO, P. et al. Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures. **Computers in Biology and Medicine**, v. 91, p. 21–37, 2017.
- CHERUBINI, A. et al. Importance of multimodal mri in characterizing brain tissue and its potential application for individual age prediction. **Journal of biomedical and health informatics**, v. 20, n. 5, p. 1232–1239, 2016.
- CHOLLET, F. et al. **Keras**. GitHub, 2015. Disponível em: <<https://github.com/fchollet/keras>>.
- COLE, J. H.; FRANKE, K. Predicting age using neuroimaging: innovative brain ageing biomarkers. **Trends in neurosciences**, v. 40, n. 12, p. 681–690, 2017.
- COLE, J. H. et al. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. **Annals of neurology**, v. 77, n. 4, p. 571–581, 2015.
- COLE, J. H. et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. **NeuroImage**, v. 163, p. 115–124, 2017.

- COLLINS, D. L. et al. Design and construction of a realistic digital brain phantom. **IEEE Transactions on Medical Imaging**, v. 17, n. 3, p. 463–468, 1998.
- COUPÉ, P. et al. Scoring by nonlocal image patch estimator for early detection of Alzheimer’s disease. **NeuroImage: clinical**, v. 1, n. 1, p. 141–152, 2012.
- COUPÉ, P. et al. Simultaneous segmentation and grading of anatomical structures for patient’s classification: application to Alzheimer’s disease. **NeuroImage**, v. 59, n. 4, p. 3736–3747, 2012.
- COUPÉ, P. et al. An optimized blockwise nonlocal means denoising filter for 3D magnetic resonance images. **IEEE transactions on medical imaging**, v. 27, n. 4, p. 425–441, 2008.
- CUINGNET, R. et al. Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database. **Neuroimage**, v. 56, n. 2, p. 766–781, 2011.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: **Computer Society Conference on Computer Vision and Pattern Recognition -CVPR**. San Diego, CA: IEEE, 2005. p. 886–893.
- DIETTERICH, T. G.; LATHROP, R. H.; LOZANO-PÉREZ, T. Solving the multiple instance problem with axis-parallel rectangles. **Artificial intelligence**, v. 89, n. 1-2, p. 31–71, 1997.
- DINSDALE, N. K. et al. Learning patterns of the ageing brain in MRI using deep convolutional networks. **NeuroImage**, v. 224, p. 117401, 2021.
- DOSIL, R.; PARDO, X. M.; FDEZ-VIDAL, X. R. Decomposition of three-dimensional medical images into visual patterns. **IEEE Transactions on Biomedical Engineering**, v. 52, n. 12, p. 2115–2118, 2005.
- DUARA, R. et al. Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. **Neurology**, v. 71, n. 24, p. 1986–1992, 2008.
- ESKILDSEN, S. F. et al. Structural imaging biomarkers of Alzheimer’s disease: predicting disease progression. **Neurobiology of Aging**, v. 36, n. 1, p. S23–S31, 2015.
- ESKILDSEN, S. F.; ØSTERGAARD, L. R. Active surface approach for extraction of the human cerebral cortex from MRI. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI**. Cham, 2006. p. 823–830.

- FELINTO, J. C. et al. Automatic segmentation and quantification of thigh tissues in CT images. In: **International Conference on Computational Science and Its Applications**. Melbourne, Australia: Springer International Publishing, 2018. p. 406–422.
- FENG, X. et al. Estimating brain age based on a healthy population with deep learning and structural mri. **Neurobiology of Aging**, v. 91, n. 1, p. 15–25, 2020.
- FERRARI, R. J.; VILLA-PINTO, C. H.; MOREIRA, C. A. F. Detection of the midsagittal plane in MR images using a sheetness measure from eigenanalysis of local 3D phase congruency responses. In: **International Conference on Image Processing (ICIP)**. Phoenix, Arizona: IEEE, 2016. p. 2335–2339.
- FERREIRA, S. I. C. **Relaxometria no estudo da doença de Parkinson**. Tese (Doutorado) — Faculdade de Ciências e Tecnologia, 2013.
- FRANKE, K. et al. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. **Neuroimage**, v. 50, n. 3, p. 883–892, 2010.
- FRASCÁ, F. N.; POLONI, K. M.; FERRARI, R. J. Automatic extraction of the midsagittal surface from T1-weighted MR brain images using a multiscale filtering approach. In: **International Conference on Computational Science and Its Applications**. Cham: Springer International Publishing, 2021. p. 131–146.
- FUJIMOTO, R. et al. Brain age estimation from T1-weighted images using effective local features. In: IEEE. **International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. Island, Korea, 2017. p. 3028–3031.
- GASER, C. et al. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer’s disease. **PloS one**, v. 8, n. 6, p. e67346, 2013.
- GASSER, T.; SROKA, L.; JENNEN-STEINMETZ, C. Residual variance and residual pattern in nonlinear regression. **Biometrika**, v. 73, n. 3, p. 625–633, 1986.
- GAUGLER, J. E. et al. Characteristics of patients misdiagnosed with Alzheimer’s disease and their medication use: an analysis of the NACC-UDS database. **BMC geriatrics**, v. 13, n. 1, p. 137, 2013.
- GOOD, C. D. et al. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. **Neuroimage**, v. 14, n. 3, p. 685–700, 2001.

- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA, USA: MIT Press, 2016. <<http://www.deeplearningbook.org>>.
- GOULDEN, C. H. **Methods of statistical analysis**. 2. ed. New York: Wiley, 1939.
- GUADALUPE, T. et al. Measurement and genetics of human subcortical and hippocampal asymmetries in large datasets. **Human Brain Mapping**, v. 35, n. 7, p. 3277–3289, 2014.
- GUDBJARTSSON, H.; PATZ, S. The Rician distribution of noisy MRI data. **Magnetic Resonance in Medicine**, v. 34, n. 6, p. 910–914, 1995.
- HAGE, M. C. F. N. S.; IWASAKI, M. Imagem por ressonância magnética: princípios básicos. **Ciência Rural**, v. 39, n. 4, p. 1275–1283, 2009.
- HALLE, M. et al. **Multi-modality MRI-based Atlas of the Brain**. 2017. <Http://www.spl.harvard.edu/publications/item/view/2037>.
- HANSON, L. G. Introduction to Magnetic Resonance Imaging techniques. Copenhagen, Denmark, p. 1–48, 2009. Disponível em: <<http://eprints.drcmr.dk/37/>>.
- HANSON, S.; PRATT, L. Comparing biases for minimal network construction with back-propagation. **Advances in neural information processing systems**, v. 1, p. 177–185, 1988.
- HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: **Fourth Alvey Vision Conference**. University of Manchester, UK: Organising Committee AVC 88, 1988. p. 10–5244.
- HE, K. et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Las Vegas, USA: IEEE, 2016. p. 770–778.
- HECKERS, S.; KONRADI, C. Hippocampal pathology in Schizophrenia. In: **Behavioral neurobiology of schizophrenia and its treatment**. Berlin, Germani: Springer, 2010. p. 529–553.
- HERBERT, M. R. et al. Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. **Brain**, v. 128, n. 1, p. 213–226, 2005.
- HERVÉ, P. Y. et al. Handedness and cerebral anatomical asymmetries in young adult males. **Neuroimage**, v. 29, n. 4, p. 1066–1079, 2006.

- HOWARD, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.
- HOZAWA, A. et al. Increased plasma 8-isoprostane levels in hypertensive subjects: the tsurugaya project. **Hypertension Research**, v. 27, n. 8, p. 557–561, 2004.
- HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: IEEE. **Conference on computer vision and pattern recognition (CVPR)**. Salt Lake City, USA, 2018. p. 7132–7141.
- HUANG, T. et al. Age estimation from brain MRI images using deep learning. In: **IEEE 14th International Symposium on Biomedical Imaging (ISBI)**. Melbourne, Australia: IEEE, 2017. p. 894–852.
- HYMAN, B. T. et al. National institute on aging - Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease. **Alzheimer’s & dementia**, v. 8, n. 1, p. 1–13, 2012.
- IGLESIAS, J. E. et al. Robust brain extraction across datasets and comparison with publicly available methods. **IEEE Transactions on Medical Imaging**, v. 30, n. 9, p. 1617–1634, 2011.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. **arXiv preprint arXiv**, abs/1502.03167, 2015.
- ITO, K. et al. Performance evaluation of age estimation from t1-weighted images using brain local features and CNN. In: IEEE. **2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. Honolulu, HI, USA, 2018. p. 694–697.
- IZENMAN, A. J. Linear discriminant analysis. In: **Modern Multivariate Statistical Techniques**. New York, NY: Springer, 2013. p. 237–280.
- JACK, C. R. J. et al. The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. **Journal of Magnetic Resonance Imaging**, v. 27, n. 4, p. 685–691, 2017.
- JAMES, B. D. et al. Contribution of Alzheimer disease to mortality in the United States. **Neurology**, v. 82, n. 12, p. 1045–1050, 2014.
- JIANG, H. et al. Predicting brain age of healthy adults based on structural MRI parcellation using convolutional neural networks. **Frontiers in neurology**, v. 10, p. 1346, 2020.

- JOHNSON, K. A. et al. Brain imaging in Alzheimer disease. **Cold Spring Harbor Perspectives in Medicine**, v. 2, n. 4, p. a006213–1–23, 2012.
- JOLLANS, L. et al. Quantifying performance of machine learning methods for neuroimaging data. **NeuroImage**, v. 199, p. 351–365, 2019.
- KAMNITSAS, K. et al. Efficient multi-scale 3D CNN with fully connected crf for accurate brain lesion segmentation. **Medical image analysis**, v. 36, p. 61–78, 2017.
- KELLEHER, J. D.; NAMEE, B. M.; D'ARCY, A. **Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies**. 1. ed. Cambridge, Massachusetts: MIT Press, 2015. 461–499 p.
- KHVOSTIKOV, A. et al. 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies. **arXiv preprint arXiv:1801.05968**, 2018.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv**, abm/1412.6980, 2014.
- KLATKA, L. A. et al. Incorrect diagnosis of Alzheimer's disease: a clinicopathologic study. **Archives of neurology**, v. 53, n. 1, p. 35–42, 1996.
- KLÖPPEL, S. et al. Automatic classification of MR scans in Alzheimer's disease. **Brain**, v. 131, n. 3, p. 681–689, 2008.
- KOROLEV, S. et al. Residual and plain convolutional neural networks for 3D brain MRI classification. In: IEEE. **International Symposium on Biomedical Imaging (ISBI)**. Melbourne, Australia, 2017. p. 835–838.
- KOVALEV, V. A.; KRUGGEL, F.; CRAMON, D. Y. von. Structural brain asymmetry as revealed by 3D texture analysis of anatomical MR images. In: **Object recognition supported by user interaction for service robots**. Quebec, Canada: IEEE, 2002. p. 808–811.
- KOVESI, P. Phase congruency: A low-level image invariant. **Psychological Research**, v. 64, n. 2, p. 136–148, 2000.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems 25**. Nevada, United States: Curran Associates, Inc., 2012. p. 1097–1105.
- LARS, S.; SVANTE, W. Analysis of variance (ANOVA). **Chemometrics and Intelligent Laboratory Systems**, v. 6, n. 4, p. 259–272, 1989.

- LECUN, Y.; BENGIO, Y. et al. Convolutional networks for images, speech, and time series. **The handbook of brain theory and neural networks**, v. 3361, n. 10, p. 1995, 1995.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.
- LI, Y. et al. Dependency criterion based brain pathological age estimation of Alzheimer’s disease patients with MR scans. **Biomedical Engineering**, v. 16, n. 1, p. 1–20, 2017.
- LIAN, C. et al. Hierarchical fully convolutional network for joint atrophy localization and Alzheimer’s disease diagnosis using structural MRI. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 4, p. 880–893, 2020.
- LIM, A. et al. Clinico-neuropathological correlation of Alzheimer’s disease in a community-based case series. **Journal of the American Geriatrics Society**, v. 47, n. 5, p. 564–569, 1999.
- LIU, M. et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer’s disease. **NeuroImage**, v. 208, n. 1, p. 116459, 2020.
- LIU, M. et al. Deep multi-task multi-channel learning for joint classification and regression of brain status. In: **International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI**. Cham: [s.n.], 2017. p. 3–11.
- LIU, M. et al. Landmark-based deep multi-instance learning for brain disease diagnosis. **Medical image analysis**, Elsevier, v. 43, p. 157–168, 2018.
- LIU, M. et al. Joint classification and regression via deep multi-task multi-channel learning for Alzheimer’s disease diagnosis. **IEEE Transactions on Biomedical Engineering**, v. 66, n. 5, p. 1195–1206, 2019.
- LIU, S. et al. Early diagnosis of Alzheimer’s disease with deep learning. In: **IEEE 11th international symposium on biomedical imaging - ISBI**. Beijing, China: IEEE, 2014. p. 1015–1018.
- LIU, S. X. Symmetry and asymmetry analysis and its implications to computer-aided diagnosis: A review of the literature. **Journal of Biomedical Informatics**, v. 42, n. 6, p. 1056–1064, 2009.

- LIU, X. et al. Symmetry identification using partial surface matching and tilt correction in 3d brain images. In: **IEEE. International Conference of the IEEE Engineering in Medicine and Biology Society**. New York, USA, 2006. p. 1056–1060.
- LIU, Y.; COLLINS, R. T.; ROTHFUS, W. E. Robust midsagittal plane extraction from normal and pathological 3-D neuroradiology images. **IEEE transactions on medical imaging**, v. 20, n. 3, p. 175–192, 2001.
- LO, C. H.; DON, H. S. 3-D moment forms: their construction and application to object identification and positioning. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 10, p. 1053–1064, 1989.
- LORENA, A. C.; CARVALHO, A. C. L. F. de. Uma introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, 2004.
- LUCARELLI, R. T. P. et al. MR imaging of hippocampal asymmetry at 3T in a multiethnic, population-based sample: results from the Dallas Heart Study. **American Journal of Neuroradiology**, v. 34, n. 4, p. 752–757, 2013.
- LUDERS, E. et al. A voxel-based approach to gray matter asymmetries. **Neuroimage**, v. 22, n. 2, p. 656–664, 2004.
- LUO, Y.; TANG, X. Automated diagnosis of Alzheimer’s disease with multi-atlas based whole brain segmentations. In: **Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging**. Bellingham, Washington: SPIE, 2017. p. 275–283.
- MADAN, C. R.; KENSINGER, E. A. Predicting age from cortical structure across the lifespan. **European Journal of Neuroscience**, v. 47, n. 5, p. 399–416, 2018.
- MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. **The annals of mathematical statistics**, v. 18, n. 1, p. 50–60, 1947.
- MARCUS, D. S. et al. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. **Journal of cognitive neuroscience**, v. 19, n. 9, p. 1498–1507, 2007.

- MARDIA, K. V.; JUPP, P. **Directional Statistics**. Washington, DC, USA: John Wiley and Sons Ltd., 2009. v. 494.
- MARON, O.; LOZANO-PÉREZ, T. A framework for multiple-instance learning. In: **Advances in neural information processing systems**. Cambridge, MA: MIT Press, 1998. p. 570–576.
- MARTÍN, A. et al. **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>.
- MARTINS, S. B. et al. Modeling normal brain asymmetry in MR images applied to anomaly detection without segmentation and data annotation. In: **Medical Imaging 2019: Computer-Aided Diagnosis**. San Diego, California: SPIE, 2019. v. 10950, p. 109500C.
- MAZZOLA, A. A. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. **Revista Brasileira de Física Médica**, v. 3, n. 1, p. 117–129, 2009.
- MCKHANN, G. M. et al. The diagnosis of dementia due to Alzheimer’s disease: recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. **Alzheimer’s & dementia**, v. 7, n. 3, p. 263–269, 2011.
- MECHANIC-HAMILTON, D. et al. Hippocampal volumetry and functional MRI of memory in temporal lobe epilepsy. **Epilepsy & Behavior**, v. 16, n. 1, p. 128–138, 2009.
- MILLER, K. L. et al. Multimodal population brain imaging in the uk biobank prospective epidemiological study. **Nature neuroscience**, v. 19, n. 11, p. 1523–1536, 2016.
- MODAT, M. et al. Fast free-form deformation using graphics processing units. **Computer Methods and Programs in Biomedicine**, v. 98, n. 3, p. 278–284, 2010.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 821. ed. Washington, DC, USA: John Wiley & Sons, 2012.
- MORRONE, M. C. et al. Mach bands are phase dependent. **Nature**, v. 324, n. 6094, p. 250–253, 1986.

- MORTENSEN, E. N.; DENG, H.; SHAPIRO, L. A SIFT descriptor with global context. In: **Computer Vision and Pattern Recognition (CVPR), 2005**. San Diego, CA, USA: IEEE, 2005. p. 184–190.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: **Proceedings of the 27th international conference on machine learning (ICML-10)**. Haifa, Israel: Omnipress, 2010. p. 807–814.
- NARASIMHAMURTHY, A. An overview of machine learning in medical image analysis: Trends in health informatics. In: **Medical imaging: Concepts, methodologies, tools, and applications**. Hershey, Pennsylvania: IGI Global, 2017. p. 36–58.
- NAVARRO, D. Learning statistics with R: A tutorial for psychology students and other beginners. University of New South Wales, New South Wales, Australia, p. 425–454, 2019.
- NELL, C.; SHAW-TAYLOR, J. **An Introduction to support vector machines and other kernel-based learning methods**. 1. ed. New York, NY, USA: Cambridge University Press, 2000. 93-124 p.
- NEWELL, A.; YANG, K.; DENG, J. Stacked hourglass networks for human pose estimation. In: **European conference on computer vision**. Amsterdam, Netherlands: Springer, 2016. p. 483–499.
- NYÚL, L. G.; UDUPA, J. K.; ZHANG, X. New variants of a method of MRI scale standardization. **IEEE Transactions on Medical Imaging**, v. 19, n. 2, p. 143–150, 2000.
- OLIVEIRA, I. A. D.; POLONI, K. M.; FERRARI, R. J. Exploring hippocampal asymmetrical features from magnetic resonance images for the classification of Alzheimer’s disease. In: **2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)**. Rochester, MN, USA: IEEE, 2020. p. 59–64.
- ORGANIZATION, W. H.; INTERNATIONAL, A. D. **Dementia: a public health priority**. 1. ed. Geneva, Switzerland: World Health Organization, 2012.
- ORTIZ, A. et al. Parkinson’s disease detection using isosurfaces-based features and Convolutional Neural Networks. **Frontiers in Neuroinformatics**, v. 13, n. 48, p. 48, 2019.
- OURSELIN, S. et al. Reconstructing a 3D structure from serial histological sections. **Image and Vision Computing**, v. 19, n. 1, p. 25–31, 2001.

- OURSELIN, S.; STEFANESCU, R.; PENNEC, X. Robust registration of multi-modal images: Towards real-time clinical applications. In: **Medical Image Computing and Computer-Assisted Intervention (MICCAI 2002)**. Heidelberg, Berlin: Springer, 2002. p. 140–147.
- PARDAKHTI, N.; SAJEDI, H. Age prediction based on brain MRI images using feature learning. In: IEEE. **International Symposium on Intelligent Systems and Informatics (SISY)**. Subotica, Servia, 2017. p. 000267–000270.
- PARDAKHTI, N.; SAJEDI, H. Brain age estimation based on 3D MRI images using 3D convolutional neural network. **Multimedia Tools and Applications**, p. 1573–7721, 2020.
- PASZKE, A. et al. Pytorch: An imperative style, high-performance deep learning library. In: WALLACH, H. et al. (Ed.). **Advances in Neural Information Processing Systems - NeurIPS**. Vancouver, Canada: Curran Associates, Inc., 2019. p. 8024–8035.
- PATENAUDE, B. et al. A bayesian model of shape and appearance for subcortical brain segmentation. **Neuroimage**, v. 56, n. 3, p. 907–922, 2011.
- PENG, H. et al. Accurate brain age prediction with lightweight deep neural networks. **Medical Image Analysis**, v. 68, p. 101871, 2021.
- PÉREZ-GARCÍA, F.; SPARKS, R.; OURSELIN, S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. **arXiv:2003.04696 [cs, eess, stat]**, mar. 2020. ArXiv: 2003.04696. Disponível em: <<http://arxiv.org/abs/2003.04696>>.
- PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. **arXiv eprint arXiv**, abs/1712.04621, 2017.
- POLONI, K. M.; FERRARI, R. J. Detection and classification of hippocampal structural changes in MR images as a biomarker for Alzheimer’s disease. In: **International Conference on Computational Science and Its Applications**. Melbourne, Australia: Springer International Publishing, 2018. p. 406–422.
- POLONI, K. M.; FERRARI, R. J. Automated detection, selection and classification of hippocampal landmark points for the diagnosis of alzheimer’s disease. **Computer Methods and Programs in Biomedicine**, p. 106581, 2021.

POLONI, K. M.; FERRARI, R. J. f. A. D. N. I. et al. A deep ensemble hippocampal cnn model for brain age estimation applied to Alzheimer's diagnosis. **Expert Systems with Applications**, Elsevier, 2022.

POLONI, K. M. et al. Brain MR image classification for alzheimer's disease diagnosis using structural hippocampal asymmetrical attributes from directional 3-d log-gabor filter responses. **Neurocomputing**, Elsevier, v. 419, p. 126–135, 2021.

POLONI, K. M. et al. Construction and application of a probabilistic atlas of 3D landmark points for initialization of hippocampus mesh models in brain MR images. In: **International Conference on Computational Science and Its Applications**. Melbourne, Australia: Springer International Publishing, 2018. p. 310–322.

QIAN, N. On the momentum term in gradient descent learning algorithms. **Neural networks**, v. 12, n. 1, p. 145–151, 1999.

REHMAN, H. Z. U.; HWANG, H.; LEE, S. Conventional and deep learning methods for skull stripping in brain MRI. **Applied Sciences**, v. 10, n. 5, p. 1773, 2020.

RESNICK, S. M. et al. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. **Journal of Neuroscience**, v. 23, n. 8, p. 3295–3301, 2003.

RUECKERT, D. et al. Nonrigid registration using free-form deformations: application to breast MR images. **IEEE Transactions on Medical Imaging**, v. 18, n. 8, p. 712–721, 1999.

SAJEDI, H.; PARDAKHTI, N. Age prediction based on brain MRI image: a survey. **Journal of medical systems**, v. 43, n. 8, p. 279, 2019.

SALVATORE, C. et al. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. **Frontiers in neuroscience**, v. 9, p. 307, 2015.

SANDLER, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: IEEE. **Conference on computer vision and pattern recognition (CVPR)**. Salt Lake City, USA, 2018. p. 4510–4520.

SARICA, A. et al. MRI asymmetry index of hippocampal subfields increases through the continuum from the Mild Cognitive Impairment to the Alzheimer's disease. **Frontiers in Neuroscience**, v. 12, n. 576, p. 1–12, 2018.

- SATO, K. et al. Neuroanatomical database of normal japanese brains. **Neural networks**, v. 16, n. 9, p. 1301–1310, 2003.
- SCHEFFE, H. **The analysis of variance**. 72. ed. Washington, DC, USA: John Wiley & Sons, Ltd., 1999.
- SCHELTENS, P. et al. Alzheimer’s disease. **The Lancet**, v. 368, n. 9533, p. 387 – 403, 2006.
- SCHELTENS, P. H. et al. Atrophy of medial temporal lobes on MRI in probable Alzheimer’s disease and normal ageing: diagnostic value and neuropsychological correlates. **Journal of Neurology, Neurosurgery & Psychiatry**, v. 55, n. 10, p. 967–972, 1992.
- SCHOLKOPF, B.; SMOLA, A. J. **Learning with kernels: support vector machines, regularization, optimization, and beyond**. 1. ed. Cambridge, Massachussets: MIT press, 2001. 187-222 p.
- SELTMAN, H. J. **Experimental design and analysis**. 1. ed. Pensilvânia, USA: Carnegie Mellow University, 2012. 171–190 p.
- SHAW, L. M. et al. Cerebrospinal fluid biomarker signature in Alzheimer ’s Disease Neuroimaging Initiative subjects. **Annals of Neurology**, v. 65, n. 4, p. 403–413, 2009.
- SHI, F. et al. Hippocampal volume and asymmetry in Mild Cognitive Impairment and Alzheimer’s disease: Meta-analyses of MRI studies. **Hippocampus**, v. 19, n. 11, p. 1055–1064, 2009.
- SIJBERS, J. et al. Maximum-likelihood estimation of Rician distribution parameters. **IEEE Transactions on Medical Imaging**, v. 17, n. 3, p. 357–361, 1998.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **CoRR**, abs/1409.1556, 2014.
- SLED, J. G.; ZIJDENBOS, A. P.; EVANS, A. C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. **IEEE Transactions on Medical Imaging**, v. 17, n. 1, p. 87–97, 1998.
- Smith, L. N. Cyclical learning rates for training neural networks. In: **IEEE. Winter Conference on Applications of Computer Vision (WACV)**. California, USA, 2017. p. 464–472.

- SØRENSEN, L. et al. Early detection of Alzheimer’s disease using MRI hippocampal texture. **Human Brain Mapping**, v. 37, n. 3, p. 1148–1161, 2016.
- SOUZA, B. da S.; POLONI, K. M.; FERRARI, R. J. Detector of 3-D salient points based on the dual-tree complex wavelet transform for the positioning of hippocampi meshes in magnetic resonance images. **Journal of Neuroscience Methods**, v. 341, p. 108789, 2020.
- SPERLING, R. Functional MRI studies of associative encoding in normal aging, Mild Cognitive Impairment, and Alzheimer’s disease. **Annals of the New York Academy of Sciences**, v. 1097, n. 1, p. 146–155, 2007.
- SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, v. 15, n. 1, p. 1929–1958, 2014.
- STEPHEN, M. S. et al. Estimation of brain age delta from brain imaging. **NeuroImage**, v. 200, p. 528–539, 2019.
- SU, L. et al. Age-related classification and prediction based on MRI: A sparse representation method. **Procedia Environmental Sciences**, v. 8, p. 645–652, 2011.
- SUTSKEVER, I. et al. On the importance of initialization and momentum in deep learning. In: **International conference on machine learning - ICML**. Atlanta, USA: JMLR.org, 2013. p. 1139–1147.
- SZEGEDY, C. et al. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition - CVPR**. Boston, USA: IEEE, 2015. p. 1–9.
- TAJBAKSH, N. et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? **IEEE transactions on medical imaging**, v. 35, n. 5, p. 1299–1312, 2016.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. **International Conference on Machine Learning**. California, USA, 2019. p. 6105–6114.
- TEVEROVSKIY, L. A. et al. Quantified brain asymmetry for age estimation of normal and AD/MCI subjects. In: **5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro**. Paris, France: IEEE, 2008. p. 1509–1512.

- THOMPSON, P. M. et al. Dynamics of gray matter loss in Alzheimer's disease. **Journal of Neuroscience**, v. 23, n. 3, p. 994–1005, 2003.
- TIELEMAN, T.; HINTON, G. **Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude**. 2012. COURSEERA: Neural Networks for Machine Learning.
- TOEWS, M. et al. Feature-based morphometry: Discovering group-related anatomical patterns. **NeuroImage**, v. 49, n. 3, p. 2318–2327, 2010.
- TOGA, A. W.; THOMPSON, P. M. Mapping brain asymmetry. **Nature Reviews Neuroscience**, v. 4, n. 1, p. 37–48, 2003.
- TUSTISON, N. J. et al. N4ITK: improved N3 bias correction. **IEEE Transactions on Medical Imaging**, v. 29, n. 6, p. 1310–1320, 2010.
- TZOURIO-MAZOYER, N. et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. **Neuroimage**, v. 15, n. 1, p. 273–289, 2002.
- UEDA, M. et al. An age estimation method using 3D-CNN from brain MRI images. In: **IEEE 16th International Symposium on Biomedical Imaging (ISBI)**. Venice, Italy: IEEE, 2019. p. 380–383.
- VAPNIK, V. N. An overview of statistical learning theory. **IEEE transactions on neural networks**, n. 5, p. 988–999, 1999.
- VEMURI, P.; JACK, C. R. Role of structural MRI in Alzheimer's disease. **Alzheimer's Research & Therapy**, v. 2, n. 4, p. 23, 2010.
- VENKATESH, S.; OWENS, R. On the classification of image features. **Pattern Recognition Letters**, v. 11, n. 5, p. 339–349, 1990.
- VICTOR, R. G. et al. The dallas heart study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. **The American journal of cardiology**, v. 93, n. 12, p. 1473–1480, 2004.
- VILLA-PINTO, C. H.; FERRARI, R. J. Initialization of deformable models in 3D Magnetic Resonance Images guided by automatically detected phase congruency point landmarks. **Pattern Recognition Letters**, v. 79, p. 1 – 7, 2016.
- VINCENT, P.; BENGIO, Y. Manifold parzen windows. In: **Advances in neural information processing systems**. Cambridge, MA: MIT Press, 2003. p. 849–856.

- WEBB, A. R. **Statistical pattern recognition**. 2. ed. Washington, DC, USA: John Wiley & Sons, Ltd., 2002. 134-141 p.
- WELCH, B. L. The generalization of student's problem when several different population variances are involved. **Biometrika**, v. 34, n. 1/2, p. 28–35, 1947.
- WIJNHOVEN, R. G.; WITH, P. de. Fast training of object detection using stochastic gradient descent. In: **Pattern Recognition (ICPR), 2010 20th International Conference on**. Istanbul, Turkey: IEEE, 2010. p. 424–427.
- WOOD, J. C.; JOHNSON, K. M. Wavelet packet denoising of magnetic resonance images: importance of Rician noise at low SNR. **Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine**, v. 41, n. 3, p. 631–635, 1999.
- WOOLARD, A. A.; HECKERS, S. Anatomical and functional correlates of human hippocampal volume asymmetry. **Psychiatry Research: Neuroimaging**, v. 201, n. 1, p. 48–53, 2012.
- YANG, C.; RANGARAJAN, A.; RANKA, S. Visual explanations from Deep 3D Convolutional Neural Networks for Alzheimer's disease classification. **arXiv preprint arXiv:1803.02544**, 2018.
- YE, T. et al. Discriminative multi-task feature selection for multi-modality classification of alzheimer's disease. **Brain imaging and behavior**, v. 10, n. 3, p. 739–749, 2016.
- YEO, I.-K.; JOHNSON, R. A. A new family of power transformations to improve normality or symmetry. **Biometrika**, v. 87, n. 4, p. 954–959, 2000.
- ZHANG, J. et al. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. **IEEE Transactions on Medical Imaging**, v. 35, n. 12, 2016.
- ZHANG, J.; LIANG, J.; ZHAO, H. Local energy pattern for texture classification using self-adaptive quantization thresholds. **IEEE transactions on image processing**, v. 22, n. 1, p. 31–42, 2012.
- ZHANG, J. et al. Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images. **IEEE Journal of Biomedical and Health Informatics**, v. 21, n. 5, p. 1607–1616, 2017.
- ZHANG, J.; LIU, M.; SHEN, D. Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. **IEEE Transactions on Image Processing**, v. 26, n. 10, 2017.

ZHANG, Y.; BRADY, M.; SMITH, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. **IEEE Transactions on Medical Imaging**, v. 20, n. 1, p. 45-57, 2001.

ZHENG, N.; XUE, J. **Statistical learning and pattern analysis for image and video processing**. 1. ed. London, New York, USA: Springer Science & Business Media, 2009. 163-170 p.