



INSTITUTO POLITÉCNICO DE BRAGANÇA Escola Superior
de Tecnologia e Gestão

Diagnóstico da Doença de Alzheimer Com Base no Electroencefalograma

Pedro Miguel de Luís Rodrigues

Relatório Final do Trabalho de Projecto apresentado à
Escola Superior de Tecnologia e de Gestão
Instituto Politécnico de Bragança

para obtenção do grau de Mestre em
Tecnologia Biomédica

[Julho de 2011]

Diagnóstico da Doença de Alzheimer Com Base no Electroencefalograma

Pedro Miguel de Luís Rodrigues

Relatório Final do Trabalho de Projecto apresentado à
Escola Superior de Tecnologia e de Gestão
Instituto Politécnico de Bragança

para obtenção do grau de Mestre em
Tecnologia Biomédica

Orientador:

Prof. Doutor João Paulo Ramos Teixeira

Supervisores de Estágio Universidade de Valladolid:

Prof. Doutor Roberto Hornero

Prof. Doutor Jesús Poza Crespo

“Este Trabalho de Projecto inclui as críticas e sugestões feitas pelo Júri”

[Julho de 2011]

*“A diferença entre o possível e o impossível
está na vontade humana.”*

Louis Pasteur

“Aos meus pais, irmão e avós”

Agradecimentos

Em primeiro lugar quero expressar a minha gratidão ao meu Orientador, Professor Dr. João Paulo Teixeira, pelo apoio, orientação, disponibilidade e motivação que me transmitiu ao longo da realização deste trabalho.

Também quero agradecer ao Grupo de Investigação de Engenharia Biomédica (Valladolid - Espanha), especialmente ao Prof. Doutor Jesús Poza e ao Prof. Doutor Roberto Hornero, pelo apoio que me deram durante os 5 meses de estágio Erasmus realizado nesse grupo, bem como aos restantes elementos pela forma como fui bem recebido por todos.

Um obrigado à Direcção Geral do Ensino Superior (DGES) e ao Instituto Politécnico de Bragança, pela Bolsa SMP-Erasmus concedida, que me possibilitou o Estágio em Espanha.

Um agradecimento especial à Unidade de Geriatria do “Hospital Universitario Pío del Río Horteiga” (Valladolid - Espanha), por me disponibilizar os sinais de EEG deste estudo.

Finalmente, um especial obrigado à minha família pelo incentivo e alento que sempre me dispensou.

O Autor,
Pedro Rodrigues

Resumo

A Doença de Alzheimer (DA) é a forma mais comum de demência. A esperança média de vida está a aumentar e esta patologia surge associada à idade, o que é preocupante. Para se alcançar um diagnóstico precoce e preciso desta doença é necessário otimizar os métodos da sua detecção. A análise de sinais do electroencefalograma (EEG) pode ser reveladora de deficiências estruturais e funcionais associadas à DA. Por conseguinte, este trabalho teve por finalidade a análise de EEGs da actividade espontânea de doentes com diagnóstico provável de DA e um grupo de controlo, usando diferentes parâmetros espectrais, bem como a criação de um sistema baseado em Redes Neurais Artificiais (RNA), de modo a discriminar sinais de EEG, entre pacientes com probabilidade de DA e sujeitos de controlo. Assim, foram aplicadas três metodologias diferentes e variações, consoante o tipo de processamento de sinal aplicado. A análise dos registos de EEGs foi realizada recorrendo à função Densidade Espectral de Potências (PSD) baseada na Transformada de Fourier de Curta Duração (STFT), à PSD baseada na Transformada de Wavelet (WT) e à aplicação da Transformada Discreta de Wavelet (DWT). A PSD baseada na WT apresentou-se neste trabalho, como uma técnica pioneira no processamento do sinal EEG e permitiu obter melhores resultados do que os conseguidos por técnicas tradicionais, como a PSD baseada na STFT e a DWT. A melhor classificação foi realizada pela RNA usando a PSD WT Biortogonal 3.5, com 0.97 de AROC, sensibilidade de 92.1%, especificidade de 90.8% e 91.5% de precisão.

Palavras-chave: Doença de Alzheimer; Electroencefalograma; Rede Neuronal Artificial; Transformada de Wavelet.

Abstract

Alzheimer's disease (AD) is the most common form of dementia. The average life expectancy is increasing and this disease arise associated with aging, which is worrying. To achieve an early and accurate diagnosis of this disease is necessary to optimize the methods of detection. The signal analysis of the electroencephalogram (EEG) may reveal structural and functional disability associated with AD. Therefore, this work aimed at the analysis of spontaneous EEGs activity of patients diagnosed with probable AD and a control group, using different spectral parameters, as well as the creation of a system based on Artificial Neural Networks (ANN) in order to discriminate the EEG signals, among patients with AD and control subjects. Three different methodologies and variations were applied, depending on the type of signal processing applied. The EEG recording analysis was performed using the Power Spectral Density (PSD) function based on Short Time Fourier Transform (STFT), the PSD based on Wavelet Transform (WT) and the application of Discrete Wavelet Transform (DWT). The PSD obtained by WT was presented in this work as a pioneer technique in the EEG signal processing and achieve better results than those obtained by traditional techniques, such as the PSD based on STFT and DWT. The best classification was performed by the ANN using PSD of WT Biorthogonal 3.5 with 0.97 of AROC , 92.1% of sensitivity, 90.8% of specificity and 91.5% of accuracy.

Keywords: Alzheimer's Disease; Electroencephalogram; Artificial Neural Networks; Wavelet Transform.

Índice

Agradecimentos	III
Resumo	IV
Abstract	V
Lista de Figuras	X
Lista de Tabelas	XIV
Índice de Acrónimos e Abreviaturas	XVII
1 Introdução	1
1.1 Motivação	1
1.2 Objectivos	1
1.3 Limitações e soluções adoptadas	2
1.4 Estado da Arte	2
1.5 Electroencefalograma	4
1.5.1 Tipos de ondas cerebrais	5
1.5.2 A actividade neuronal do cérebro humano	6
1.6 Doença de Alzheimer	8
1.6.1 Factores de risco	10
1.6.2 Sintomas da doença de Alzheimer	10
1.6.3 Diagnóstico	11
1.6.4 Tratamento	11
1.6.5 Diferenças no sinal de EEG de um paciente com DA em relação a um sujeito de controlo	12

1.7	Organização da Dissertação	13
2	Processamento do Sinal de EEG	14
2.1	Sinal Biomédico	14
2.1.1	Ruídos possíveis em Sinais Biomédicos	15
2.1.1.1	Ruídos devido aos instrumentos utilizados	15
2.1.1.2	Ruídos devido ao ambiente de experimentação	15
2.2	Transformada de Fourier	15
2.3	Transformada de Fourier de Curta Duração	17
2.4	Transformada Wavelet	18
2.5	Normalização da Densidade Espectral de Potência	21
2.6	Seleção de características espectrais	21
3	Métodos de Classificação	24
3.1	Redes Neurais Artificiais	24
3.1.1	Tipo de arquiteturas para Redes Neurais Artificiais mais utilizadas	25
3.1.1.1	Redes Feed-forward	25
3.1.1.1.1	Redes Feed-forward com multicamadas	26
3.1.2	Redes Neurais Recorrentes	26
3.1.3	Métodos de aprendizagem das redes neuronais artificiais	27
3.1.3.1	Aprendizagem supervisionada	27
3.1.3.1.1	Processamento Neuronal numa aprendizagem supervisionada	27
3.1.3.2	Aprendizagem não supervisionada	27
3.1.4	Função de activação de uma Rede Neuronal Artificial	28
3.1.5	Métodos de Treino das redes neuronais artificiais Supervisionadas	28
3.1.5.1	Algoritmo Back-Propagation	28
3.1.5.2	Algoritmo do Gradiente Conjugado - <i>Scaled Conjugate Gradient Algorithm</i>	29
3.1.6	Função de erro cross-entropy para duas classes	30
3.1.7	Weight Decay	32
3.1.8	Registo cronológico do desenvolvimento das redes neuronais	32
3.1.9	Aplicações das Redes Neurais Artificiais	33
3.1.9.1	Condições para a aplicação da RNA	33
3.1.10	Validação cruzada deixando um de fora (leave - one - out - cross - validation)	34
3.2	Regressão Logística	34

3.2.1	Regressão Logística Binária	34
3.3	Métodos de apoio à decisão	35
3.3.1	Curvas ROC	35
3.3.2	<i>P</i> -Valores	37
3.3.2.1	Teste de Kolmogorov-Smirnov	39
3.3.2.2	Teste de Levene	39
3.3.2.3	Teste de Kruskal-Wallis	39
3.3.3	Análise de componentes principais (PCA)	40
4	Classificação dos Sinais EEG	42
4.1	Recolha e Pré-Processamento do Sinal de EEG	42
4.1.1	Seleccção dos pacientes e dos sujeitos de controlo	42
4.1.2	Gravação dos Sinais de EEG	42
4.1.3	Pré-Processamento dos sinais de EEG	43
4.2	Classificação dos sinais EEG utilizando a <i>PSD</i> baseada na WT	44
4.2.1	Processamento do sinal EEG e extracção de características	44
4.2.2	Utilização dos parâmetros <i>RP</i> e rácios espectrais	45
4.2.2.1	Análise estatística / selecção de características	45
4.2.2.2	Parâmetros <i>RP</i> e rácios espectrais por eléctrodo	50
4.2.2.2.1	Classificação dos sinais EEG	51
4.2.2.3	Parâmetros <i>RP</i> e rácios espectrais - “ <i>grand average</i> ” de 5s	53
4.2.2.3.1	Classificação dos sinais EEG	53
4.2.3	Utilização de alguns parâmetros espectrais	55
4.2.3.1	Análise estatística / selecção de características	55
4.2.3.2	Parâmetros espectrais por eléctrodo	58
4.2.3.2.1	Classificação dos sinais EEG	59
4.2.3.3	Parâmetros espectrais - “ <i>grand average</i> ” de 5s	61
4.2.3.3.1	Classificação dos sinais EEG	61
4.3	Classificação dos sinais EEG utilizando a <i>PSD</i> baseada na STFT	63
4.3.1	Processamento do sinal EEG e extracção de características	63
4.3.2	Utilização dos parâmetros <i>RP</i> e rácios espectrais	63
4.3.2.1	Análise estatística / selecção de características	63
4.3.2.2	Parâmetros <i>RP</i> e rácios espectrais por eléctrodo	67
4.3.2.2.1	Classificação dos sinais EEG	67

4.3.2.3	Parâmetros <i>RP</i> e rácios espectrais - “ <i>grand average</i> ” de 5s	69
4.3.2.3.1	Classificação dos sinais EEG	69
4.3.3	Utilização de alguns parâmetros espectrais	71
4.3.3.1	Análise estatística / selecção de características	71
4.3.3.2	Parâmetros espectrais por eléctrodo	74
4.3.3.2.1	Classificação dos sinais EEG	74
4.3.3.3	Parâmetros espectrais - “ <i>grand average</i> ” de 5s	76
4.3.3.3.1	Classificação dos sinais EEG	76
4.4	Classificação dos sinais EEG com recurso à DWT	78
4.4.1	Análise estatística / selecção de características	79
5	Discussão dos resultados	86
5.1	Aspectos gerais	86
6	Conclusão	90
	Referências	92

Lista de Figuras

1.1	Distribuição dos eléctrodos e dos potenciais eléctricos no cérebro humano.	4
1.2	Ondas cerebrais delta, teta, alfa, beta e gama.	6
1.3	Estrutura de um neurónio.	7
1.4	O potencial de acção viaja ao longo do axónio sem perdas de sinal.	7
1.5	Placas de Proteínas β -amiloide.	9
1.6	Emaranhados neurofibrilares.	9
1.7	Estágios: Leve, Moderado e Grave da Doença de Alzheimer.	11
1.8	Espectro característico de um paciente com Alzheimer e de um sujeito de controlo.	12
2.1	Exemplo de Árvore de decomposição da WPT até ao nível 3.	20
3.1	Neurónio Artificial.	25
3.2	Rede Neuronal Feed-forward.	25
3.3	Redes Neurais tipo Feed-Forward com múltiplas camadas.	26
3.4	Rede Neuronal Recorrente.	26
3.5	Diagrama do processo de aprendizagem supervisionada.	27
3.6	Processo de validação cruzada deixando um de fora.	34
3.7	Curva Roc.	36
4.1	Processo de eliminação da média por eléctrodo.	43
4.2	5 segundos de Sinal EEG.	43
4.3	5 segundos de Sinal EEG filtrado.	44
4.4	Exemplo de um Processo de cálculo da média das PSD_n de 5s de duração (PSD_n por eléctrodo).	45

4.5	Número de canais significativos para os processamentos envolvidos neste estudo com p -valores <0.05 obtidos através do teste de Kruskal-Wallis com correcção de Bonferroni.	46
4.6	Diagramas de caixa que apresentam a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais obtidos pela $PSD WT$ Bior 3.5	47
4.7	Diagramas de caixa que mostram a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais obtidos pela $PSD WT$ Bior 3.5 - Continuação.	48
4.8	Cabeças que representam as diferenças entre os pacientes de Alzheimer e os sujeitos de controlo por eléctrodo com um $p<0.05$	50
4.9	Biplot das componentes principais das características $r_1, r_2, RP(beta1), RP(teta)$	51
4.10	Valores de $AROC$ ao longo dos nós da camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	52
4.11	Biplot das componentes principais das características $r_1, r_2, RP(beta1), RP(teta)$	53
4.12	Valores de $AROC$ ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	54
4.13	Número de canais significativos para os processamentos envolvidos neste estudo com p -valores <0.05 obtidos através do teste de Kruskal-Wallis com correcção de Bonferroni.	56
4.14	Diagramas de caixa que mostram a distribuição dos valores de alguns parâmetros espectrais pela $WT COIF 3$	57
4.15	Cabeças que representam as diferenças entre os pacientes de Alzheimer e os sujeitos de controlo por eléctrodo com um $p<0.05$	58
4.16	Biplot das componentes principais das características SE, IAF, MF	59
4.17	Valores de $AROC$ ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	60
4.18	Biplot das componentes principais das características SE, IAF, MF	61
4.19	Valores de $AROC$ ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	62
4.20	Diagramas de caixa que mostram a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais por indivíduo, obtidos pela $PSD STFT$.	64

4.21	Diagramas de caixa que mostram a distribuição dos valores de <i>RP</i> em cada faixa de frequência clássica e respectivos rácios espectrais por indivíduo, obtidos pela <i>PSD STFT</i> - continuação.	65
4.22	Cabeças que representam as diferenças entre os pacientes de DA e os sujeitos de controlo por eléctrodo com um $p < 0.05$	67
4.23	Biplot das componentes principais das características $r_1, r_2, RP(beta1), RP(teta)$	68
4.24	Valores de <i>AROC</i> ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	68
4.25	Biplot das componentes principais das características $r_1, r_2, RP(beta1), RP(teta)$	70
4.26	Valores de <i>AROC</i> ao longo dos nós da camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	70
4.27	Diagramas de caixa que mostram a distribuição dos valores de alguns parâmetros espectrais e correspondentes <i>p</i> -valores para todos os canais obtidos pela <i>STFT</i>	72
4.28	Cabeças que representam as diferenças entre os pacientes com AD e os sujeitos de controlo por eléctrodo com um $p < 0.05$	73
4.29	Biplot das componentes principais das características <i>SE, IAF, MF</i>	74
4.30	Valores de <i>AROC</i> ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	75
4.31	Biplot das componentes principais das características <i>SE, IAF, MF</i>	77
4.32	Valores de <i>AROC</i> ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	77
4.33	Decomposição do sinal pela WT até a decomposição 5	79
4.34	Número de canais significativos para os Processamentos envolvidos neste estudo com p -valores < 0.05 obtidos através do teste de Kruskal-Wallis com correcção de Bonferroni.	80
4.35	Diagramas de caixa que mostram a distribuição dos valores de <i>RP</i> em cada faixa de frequência clássica e respectivos rácios espectrais pela WT Bior 3.5.	81
4.36	Diagramas de caixa que mostram a distribuição dos valores de <i>RP</i> em cada faixa de frequência clássica e respectivos rácios espectrais pela WT Bior 3.5 - Continuação.	82
4.37	Cabeças que representam as diferenças entre os pacientes de Alzheimer e os sujeitos de controlo, por eléctrodo, com um $p < 0.05$	83

4.38	Biplot das componentes principais das características $e_1, e_2, RP(teta)$	83
4.39	Valores de $AROC$ ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.	84

Lista de Tabelas

1.1	Principais gamas de Frequência do sinal EEG.	6
2.1	Propriedades das Famílias mais importantes de Wavelet	20
3.1	Funções de activação.	28
3.2	Matriz de Confusão / tabela de contingência.	37
3.3	Tomada de decisão de um teste de Hipóteses.	38
4.1	Teste de KS e teste de Levene das distribuições das RP s e dos rácios espectrais - por indivíduo.	45
4.2	Médias e desvios padrão dos diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-Wallis, de parâmetros extraídos da PSD_n baseada na WT Bior 3.5, por indivíduo.	49
4.3	Resultados de discriminação, obtidos através da curva ROC, dos parâmetros extraídos da PSD_n baseada na WT Bior 3.5, por indivíduo.	49
4.4	Conjuntos de Treino e de Teste.	51
4.5	Resultados da classificação do grupo de teste.	52
4.6	Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.	53
4.7	Conjuntos de Treino e de Teste.	54
4.8	Resultados da classificação do grupo de teste.	55
4.9	Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.	55
4.10	Teste de KS e teste de Levene das distribuições das RP s e dos rácios espectrais, obtidos por indivíduo	56
4.11	Médias e desvios padrão dos Diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-Wallis, de parâmetros espectrais extraídos da PSD_n baseada na WT Coif. 3, por indivíduo.	58

4.12 Resultados de discriminação obtidos através da curva ROC, dos parâmetros espectrais extraídos da PSD_n baseada na WT Coif. 3, por indivíduo.	58
4.13 Conjuntos de Treino e de Teste.	59
4.14 Resultados da classificação do grupo de teste.	60
4.15 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	60
4.16 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	62
4.17 Resultados da classificação do grupo de teste.	62
4.18 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	63
4.19 Teste de Ks e teste de Levene das distribuições das RPs e dos rácios espectrais - por indivíduo	64
4.20 Médias e desvios padrão dos Diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-Wallis, de parâmetros extraídos da PSD_n baseada na STFT, por indivíduo. . . .	66
4.21 Resultados de discriminação, obtidos através da curva ROC, dos parâmetros extraídos da PSD_n baseada STFT, por indivíduo.	66
4.22 Conjuntos de Treino e de Teste.	68
4.23 Resultados da classificação do grupo de teste	69
4.24 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	69
4.25 Conjuntos de Treino e de Teste.	70
4.26 Resultados da classificação do grupo de teste.	71
4.27 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	71
4.28 Teste de KS e teste de Levene das distribuições das RPs e dos rácios espectrais.	72
4.29 Médias e desvios padrão dos Diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-Wallis para a PSD_n da STFT.	73
4.30 Resultados de discriminação obtidos através da curva ROC.	73
4.31 Conjuntos de Treino e de Teste.	75
4.32 Resultados da classificação do grupo de teste	76
4.33 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	76
4.34 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	77
4.35 Resultados da classificação do grupo de teste.	78
4.36 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	78
4.37 Teste de KS e teste de Levene das distribuições das RPs e dos rácios espectrais - por indivíduo.	79
4.38 Médias e desvios padrão dos Diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-wallis - por eléctrodo.	82

4.39	Resultados de discriminação obtidos através da curva ROC, por indivíduo	82
4.40	Conjuntos de Treino e de Teste.	84
4.41	Resultados da classificação do grupo de teste.	85
4.42	Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos. . .	85
5.1	Tabela resumo dos resultados de classificação das Redes Neurais obtidos nas três Metodologias, utilizando as <i>RPs/ERs</i> e rácios espectrais.	88
5.2	Tabela resumo dos resultados de classificação das Redes Neurais obtidos nas duas metodologias baseadas na <i>PSD</i> , utilizando alguns parâmetros espectrais.	88

Índice de Acrónimos e Abreviaturas

ACh	Acetilcolina
APOE	Apolipoproteína E
AROC	Área a baixo da curva ROC
BD	Base de dados
Bior	<i>Biortogonal</i>
Coif	<i>Coiflet</i>
CWD	Distribuição de Choi-Williams
DA	Doença de Alzheimer
Db	<i>Daubechies</i>
DTFT	Transformada discreta de Fourier
DWT	Transformada Discreta de Wavelet
e_1	Rácio espectral DWT 1
e_2	Rácio espectral DWT 2
e_3	Rácio espectral DWT 3
ECG	Electrocardiograma
ECoG	Registos eléctricos corticais
EEG	Electroencefalograma
EEGq	Electroencefalograma Quantitativo
EMG	Electromiograma
ER	Percentagem de Energia
$ER(\alpha)$	Percentagem de Energia em alfa
$ER(\beta)$	Percentagem de Energia em beta
$ER(\delta)$	Percentagem de Energia em delta
$ER(\gamma)$	Percentagem de Energia em gama
$ER(\theta)$	Percentagem de Energia em teta

<i>IAF</i>	Frequência individual de alfa
KS	Kolmogorov-Smirnov
MEG	Magnetoencefalograma
<i>MF</i>	Frequência média
N.S.	Não significativa
<i>p</i>	<i>p</i> -valor
PCA	Análise de componentes principais
PET	Tomografia por Emissão de Positrão
<i>PSD</i>	Densidade Espectral de Potência
r_1	Rácio espectral 1
r_2	Rácio espectral 2
r_3	Rácio espectral 3
r_4	Rácio espectral 4
REM	<i>Rapid Eye Movement</i>
RL	Regressão Logística
RMN	Ressonância Magnética Nuclear
RNA	Rede Neuronal Artificial
ROC	<i>Receiver Operating Characteristic</i>
<i>RP</i>	Potência relativa
<i>RP(alfa)</i>	Potência relativa em alfa
<i>RP(beta1)</i>	Potência relativa em beta1
<i>RP(beta2)</i>	Potência relativa em beta2
<i>RP(delta)</i>	Potência relativa em delta
<i>RP(gama)</i>	Potência relativa em gama
<i>RP(teta)</i>	Potência relativa em teta
RTF	Representações Tempo-Frequência
SD	Desvio padrão
<i>SE</i>	Entropia Espectral
<i>SEF95%</i>	Frequência de fronteira espectral 95%
SPECT	Tomografia Computadorizada de Perfusão por Emissão de Fóton
STFT	Transformada de Fourier de curta duração
Sym	<i>Symlet</i>
TC	Tomografia Computorizada
WD	<i>Weight decay</i>

WPT	<i>Wavelet Packet Transform</i>
WT	Transformada de Wavelet
WVD	Distribuição de Wigner-Ville

Introdução

Este capítulo constitui uma nota introdutória aos estudos que são abordados nesta Dissertação, apresentando os objectivos, o estado da arte, a motivação da realização, as limitações e soluções adoptadas e a forma como se encontra organizada. Realiza também a descrição do electroencefalograma e da actividade neuronal. Finalmente, o capítulo encerra com a relevância actual da Doença de Alzheimer, explicando a sua sintomatologia, evolução, técnicas de diagnóstico e tratamento.

1.1 Motivação

A motivação da pesquisa desenvolvida surge porque a DA é a forma mais comum de demência no mundo ocidental, onde é esperado um grande aumento da sua prevalência devido ao aumento da esperança média de vida. Por outro lado, a precisão de diagnóstico é relativamente baixa e não existe ainda um biomarcador capaz de detectar definitivamente esta demência sem testes invasivos [1, 2]. Como se trata de uma demência cortical, o electroencefalograma (EEG) pode ser revelador de deficiências estruturais e funcionais associadas a esta doença, bem como constituir um exame complementar de diagnóstico capaz de possibilitar um diagnóstico diferencial da DA, cada vez mais atempado, eficiente, preciso e correcto. Aliás, o estudo dos registos electroencefalográficos são aqueles que melhor ajudam a entender a neuropatologia da DA.

Por tudo isto, julga-se pertinente desenvolver novas técnicas que ajudem a melhorar o diagnóstico da DA.

1.2 Objectivos

O Objectivo fundamental desta Dissertação consiste em desenvolver um Sistema Artificial, baseado nas Redes Neurais Artificiais (RNA), que permita discriminar sinais de Electroencefalograma entre pacientes com probabilidade de terem DA e sujeitos de controlo.

Para a consecução do objectivo geral, delinearam-se os seguintes objectivos específicos:

1. Obtenção de uma base de dados de Sinais de EEG de pacientes com DA e sujeitos de controlo;
2. Pesquisa bibliográfica de métodos de análise tempo-frequência, aplicados aos sinais Biomédicos em geral e ao EEG em particular;
3. Análise e discussão dos resultados obtidos;
4. Publicação dos resultados mais importantes em conferências nacionais e internacionais, e em revistas internacionais de referência.

1.3 Limitações e soluções adoptadas

Toda a pesquisa científica impõe dificuldades que não podem ser desconsideradas e omitidas pelo investigador. A principal dificuldade encontrada, foi a obtenção da Base de Dados (BD), fundamental para a realização deste estudo.

Numa tentativa de superar essa dificuldade, o Grupo de Investigação Biomédica de Valladolid (Espanha) em parceria com a Unidade de Geriatria do “Hospital Universitario Pío del Río Hortega” - Valladolid (Espanha) facultou a BD necessária à realização deste trabalho.

1.4 Estado da Arte

Os primeiros trabalhos que visaram atingir transformações não lineares foram realizados por PE Wigner (1932), J. Ville (1948) e J.E. Moyal (1949) no campo da mecânica quântica. Simultaneamente, D. Gabor lançou as bases para a maioria das representações tempo-frequência, conhecidas até hoje [3]. A partir daí, a investigação centrou-se nas Representações Tempo-frequência (RTF), C.H. Page (1952), H. Margenau e R.N. Hill (1961), W. Rihaczek (1968), H.I. Choi e W.J. Williams (1989) e J. Jeong e W.J. Williams (1992). No entanto, foi L. Cohen, quem utilizou o conceito de Função Kernel e Teoria de Operadores para derivar uma classe geral da RTF [4]. Permitindo assim desenvolver muitas das RTFs, que surgiram até aos nossos dias.

A importância da RTF é evidenciada pela sua crescente aplicação em diversos tipos de sinais não-estacionários fisiológicos [5]. Por exemplo, a WT foi aplicada para detectar o complexo QRS do electrocardiograma (ECG) [6]. Outros trabalhos compararam RTFs diferentes como o espectrograma e a WT para discriminar entre ECGs de pacientes suscetíveis à ocorrência de arritmias e ECGs de sujeitos de controlo (Jones 1992). A STFT e a WT também foram avaliadas quanto à sua capacidade de caracterizar fonocardiogramas anormais [7]. Por outro lado, a Distribuição de Choi-Williams (CWD) tem sido usada para estudar os ritmos anormais no electrogastrogramas [8]. Estes autores também aplicaram a WT e a

Distribuição de Wigner-Ville (WVD) para estudar a resposta dos neurónios auditivos após estimulação com ruído de banda larga [9]. Para o estudo dos registos de actividade cerebral Zaveri (1992) usou o espectrograma, a WVD e a CWD para caracterizar o comportamento estacionário de registos eléctricos corticais (ECoG) de pacientes epiléticos. A WT também tem sido utilizada para identificar o aparecimento de crises epiléticas no EEG [10], para caracterizar mudanças no potencial evocado e para discriminar EEGs e Magnetoencefalogramas (MEG) de pacientes com AD [11, 12].

A análise espectral do EEG foi um método utilizado pela primeira vez em 1932 por Dietsch sendo estudada posteriormente por diversos autores como um exame complementar de diagnóstico.

Partindo destas ideias, no presente trabalho desenvolver-se-à a investigação, extraindo uma série de parâmetros do sinal de EEG após processamento por várias RTFs baseadas na STFT e na WT, de modo a caracterizar a actividade espontânea do EEG nos pacientes com um provável diagnóstico de DA e sujeitos de controlo. Pretende-se também avaliar a capacidade da WT face aos métodos espectrais clássicos.

Como a DA é uma demência cortical, é razoável supor que as anomalias observadas na actividade cerebral, revelem défices anatómicos e funcionais nessa parte do cérebro. Assim, desenvolveram-se diversos estudos com uma análise exaustiva acerca dos padrões da actividade base do EEG na DA. Uma das anomalias mais observadas na actividade do EEG de pacientes de DA consiste numa desaceleração dos ritmos observados. A DA está associada a um aumento da potência nas bandas de baixa frequência (delta e teta) e uma diminuição da potência nas frequências mais altas (alfa e beta). Outros estudos explicam a desaceleração do EEG pela perda do neurotransmissor acetilcolina, verificaram que após a administração de medicamentos colinérgicos a memória de pacientes portadores de DA melhorou, e os sinais de EEG dos mesmos pacientes apresentaram uma tendência de mudança para um padrão normal. De igual modo, constataram que após a administração de medicamentos anti-colinérgicos se registava uma desaceleração do sinal de EEG [13]. Um grande número de estudos sugere que as anomalias registadas nas conexões sinápticas podem esconder muitas alterações neurológicas, como a DA, Parkinson e Epilepsia. Outros estudos desenvolvidos sugerem que a coerência da actividade do EEG pode ajudar a compreender a associação entre a plasticidade sináptica e o rendimento cognitivo [14]. De igual modo, do ponto de vista físico, investigou-se a dinâmica não linear do EEG na DA para compreender a sua importância nas funções cerebrais. Nesse sentido, a análise dos ritmos de EEG mostrou uma diminuição da complexidade e da conectividade funcional da DA, devido a uma redução e a uma dinâmica não linear da actividade neuronal e/ou nas conexões corticais.

Face ao descrito, neste trabalho procurar-se-á verificar a desaceleração do EEG dos pacientes com probabilidade de DA, de forma a extrair parâmetros do sinal que permitam discriminar os Sinais de EEG entre pacientes com DA e sujeitos de controlo.

Constituindo métodos válidos para a solução de problemas complexos de difícil resolução, através

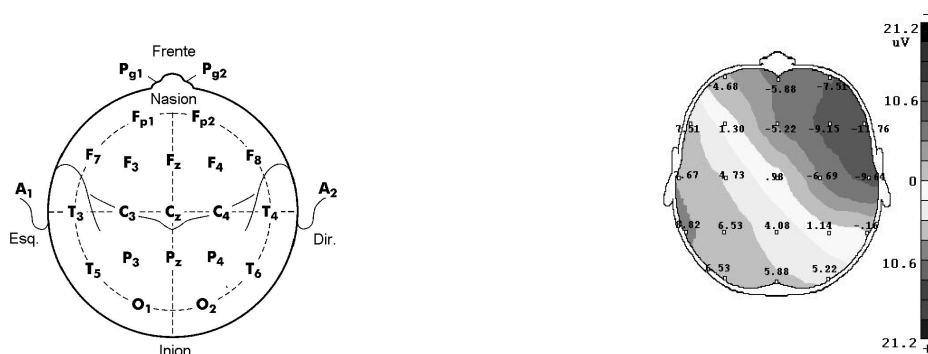
de inteligência artificial, as RNAs têm vindo a assumir, cada vez mais, o papel de uma valiosa ferramenta para diversas aplicações. No que respeita à utilização das RNAs para a detecção da DA, foram realizados estudos sobre a detecção precoce da DA por Modelação Bump (processamento baseado na WT Morlet) de sinais de EEG por Vialatte em 2005. Este método é relativamente barato face aos métodos tradicionais para detectar a doença precocemente, apresentou excelentes resultados pois teve uma precisão na ordem dos 80% a 93% [15]. Outro estudo surgiu no mesmo ano, sobre a classificação de sinais de EEG utilizando RNAs e Regressão logística (RL) na detecção de Epilepsia, por Abdulhamit Subasi [16]. Este estudo mostrou que pode vir a ser útil no diagnóstico da epilepsia (precisão a rondar os 90%). Investigações sobre a classificação de sinais de EEG utilizando a energia relativa de Wavelet e RNAs foram efectuadas por Ling Guo em 2009, para a detecção da epilepsia. Os resultados foram muito bons e a RNA teve um excelente desempenho (precisão de 95%) [17].

No presente estudo as RNAs e a RL vão ter um papel relevante na discriminação dos sinais EEG entre pacientes com DA e sujeitos de controlo.

1.5 Electroencefalograma

O EEG é uma técnica de exame cerebral usada desde 1929, depois do psiquiatra alemão Hans Berger descobrir que o cérebro gerava uma actividade eléctrica capaz de ser registada. O EEG tem-se revelado desde então, uma ferramenta muito útil no estudo do estado funcional e no diagnóstico de danos e distúrbios do cérebro, é muito utilizado por ser um exame de diagnóstico sensível, não invasivo, economicamente acessível e de elevada resolução temporal [18].

Figura 1.1 Distribuição dos eléctrodos e dos potenciais eléctricos no cérebro humano.



Fonte: Adaptado de [19]

Para se realizar esta técnica de exame, procede-se à colocação de eléctrodos no couro cabeludo do paciente utilizando uma pasta electrolítica (o interface eléctrodo-electrólito é muito importante e deve ser capaz de fornecer uma boa conexão eléctrica entre a pele e o material condutor do eléctrodo, para

evitar distorções do sinal de EEG). A posição de fixação dos eléctrodos é feita de acordo com um padrão internacional (Sistema Internacional 10-20) que determina a posição específica para cada eléctrodo [20]. Na Figura 1.1 pode observar-se a distribuição dos electrodos no Lóbulo Frontal - “F”, Lóbulo Temporal - “T”, Lóbulo Central - “C”; Lóbulo Parietal - “P” e Lóbulo Occipital- “O”. Há a referir que não existe lóbulo central no córtex cerebral. Por isso, “C” é usado apenas para a identificação proposta na figura. Os números (2, 4, 6, 8) são relativos ao hemisfério direito, os restantes números (1, 3, 5, 7) são relativos ao hemisfério esquerdo. “Z” refere-se ao eléctrodo que está posicionado no centro cerebral. “Fp” significa Frente Polar. “Nasion” é o ponto entre a testa e o nariz. “Inion” é o ponto mais saliente da parte occipital do cérebro. Os eléctrodos são depois ligados a um amplificador que aumenta a amplitude do sinal eléctrico recolhido pelos eléctrodos milhares de vezes. Esse sinal é convertido num sinal digital através de um sistema de aquisição de dados, para posterior visualização, armazenamento, processamento e análise num computador. Há dois tipos de exame de EEG: os registos espontâneos, onde as diferenças de potencial são medidas continuamente e sem a presença de estímulos exteriores e os registos evocados que são os potenciais associados à resposta cerebral a um estímulo que pode ser auditivo, visual ou sensorial. Os potenciais evocados apresentam amplitude muito baixa (1-30 μV) relativamente à actividade do EEG espontâneo. As características do traçado dependem consideravelmente do estado de repouso do indivíduo[20].

1.5.1 Tipos de ondas cerebrais

As ondas cerebrais são formas de ondas electromagnéticas produzidas pela actividade eléctrica das células cerebrais. Podem ser medidas com aparelhos electrónicos como o electroencefalograma (EEG). As frequências dessas ondas eléctricas são medidas em ciclos por segundo ou Hertz (Hz) [21].

As ondas cerebrais mudam de frequência baseadas na actividade eléctrica dos neurónios e estão relacionadas com mudanças de estados de consciência. Cada ser humano tem características próprias na sua actividade cerebral. Essa actividade tem um padrão e um ritmo, e incorpora as frequências Delta, Teta, Alfa, Beta, e Gama, em vários níveis (Tabela 1.1 e Figura 1.2).

A Delta é a mais baixa de todas as frequências de ondas cerebrais. Está associada ao sono profundo. A faixa Delta situa-se entre 0.5 e 4 Hz [22].

A banda de frequência de 4 a 8 Hz é a banda Teta. É a banda de frequência predominante durante o sono leve de adultos e tem uma amplitude menor que as outras bandas. O movimento rápido dos olhos, que é um comportamento característico ao sonhar, designado REM - *Rapid Eye Movement* está na banda Teta [22].

A banda Alfa situa-se aproximadamente entre 8 e 13 Hz, sendo de maior magnitude e predominante na região posterior do córtex. Esta banda está associada ao estado de vigília em repouso, estando a pessoa

acordada de olhos fechados [23].

A faixa de ondas Beta situa-se entre 13 e 30 Hz. O estado Beta está essencialmente associado à concentração, à atenção redobrada, à melhor acuidade visual e à coordenação.

A onda Gama é definida como um padrão das ondas cerebrais associado à percepção e à consciência. As ondas Gama são produzidas quando massas de neurónios emitem sinais eléctricos à taxa aproximada de quarenta vezes por segundo e ocorre a partir de frequências superiores a 30 Hz [21].

Figura 1.2 Ondas cerebrais delta, teta, alfa, beta e gama.

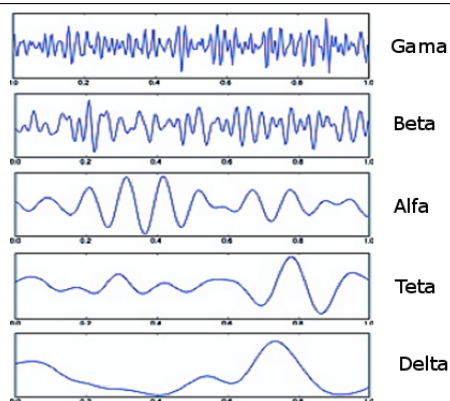


Tabela 1.1 Principais gamas de Frequência do sinal EEG.

Tipos de onda	Frequências (Hz)
Delta	0.5-4
Teta	4-8
Alfa	8-13
Beta	13-30
Gama	>30

1.5.2 A actividade neuronal do cérebro humano

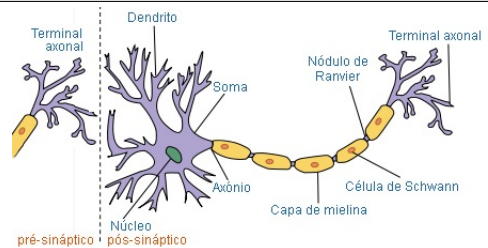
O sistema nervoso é constituído por uma complexa rede de interligação entre os vários tecidos nervosos, está envolvido em tarefas de controlo e integração das actividades do corpo e nele são gerados uma série de sinais eléctricos que dão lugar aos registos de EEG. A unidade básica do sistema nervoso central é a célula nervosa, chamada neurónio, que é constituído pelo corpo celular ou soma (onde se encontra o núcleo celular), dendrites (são extensões bastante curtas do corpo celular e estão envolvidas na recepção de estímulos) e o axónio (é geralmente uma única extensão alongada) [24].

O neurónio electricamente excitável é capaz de gerar um potencial de acção que corre rapidamente pelo seu prolongamento. Quando chega ao fim desse prolongamento ocorre a libertação de neurotransmissores que excitam outra célula, conclui desse modo uma transmissão de sinais, a chamada

sinapse[24].

O neurónio pré-sináptico armazena e liberta o neurotransmissor; o neurónio pós-sináptico contém os receptores do neurotransmissor; a fenda sináptica é o espaço entre neurónios onde é libertado o neurotransmissor; a transmissão sináptica ocorre quando se dá a união do neurotransmissor pré-sináptico com o seu receptor no neurónio pós-sináptico (Conf. Figura 1.3).

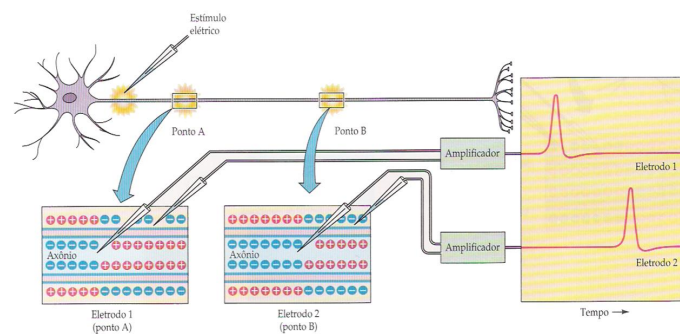
Figura 1.3 Estrutura de um neurónio.



Fonte: Adaptado de [25]

A maior parte das sinapses que existem no organismo humano, são sinapses de tipo químico. Os sinais químicos que passam dos terminais axonais para as dendrites da próxima célula são transformados em sinais eléctricos. Estes sinais são integrados, reforçando ou inibindo os sinais eléctricos recebidos por todas as outras sinapses do neurónio e deste processo resulta a decisão final de enviar (ou não) o sinal através do axónio para a célula seguinte. Os potenciais eléctricos gerados são enviados pelo axónio até às próximas sinapses nas dendrites do neurónio seguinte, onde o processo é repetido [26]. Existe sempre uma diferença de potencial ou tensão eléctrica entre o interior e o exterior de uma célula. Isso acontece devido a uma desigual distribuição de iões entre os dois lados da membrana e à permeabilidade da membrana a esses iões. A tensão de uma célula inactiva permanece num valor negativo (considerando o interior da célula em relação ao exterior). Quando a membrana de uma célula excitável é despolarizada além do limiar de -55mV a célula dispara um potencial de acção, de amplitude sempre igual. O potencial de acção é o meio através do qual a informação é transportada ao longo do axónio (Ver Figura 1.4).

Figura 1.4 O potencial de acção viaja ao longo do axónio sem perdas de sinal.



Fonte: Adaptado de [20]

Todo o sistema nervoso pode gerar actividade eléctrica. No córtex cerebral as células são organizadas hierarquicamente em seis camadas de acordo com Korbinian Brodmann [27]: Camada I ou Camada Molecular, é pobre em número de células; Camada II ou Granular Externa, composta por células granulares; Camada III ou Piramidal Externa, composta por células piramidais; Camada IV ou Granular Interna, composta por células granulares; Camada V ou Piramidal Interna, composta por células piramidais; Camada VI ou Multiforme, contém diferentes tipos de células. Dado que cerca de 70% das células no neo-córtex são piramidais com dendrites apicais que se estendem desde o soma até à superfície do córtex, as correntes extracelulares extraídas através dos eléctrodos no escalpo são o resultado de centenas de milhares de neurónios piramidais [28]. O tecido cortical produz actividade eléctrica espontaneamente, gera ondas lentas com ritmos rápidos sobrepostos. Todavia, para que estes sinais ocorram deve existir actividade sináptica em pelo menos uns 6 cm de córtex cerebral. Estas áreas correspondem a três zonas de geração cortical e são designadas geradores. O gerador A está localizado a cerca de 500 μm da superfície cortical e ao despolarizar-se origina ondas negativas na superfície das dendrites apicais das células piramidais. O Gerador B é mais interior do que o anterior, situando-se a 900 μm da superfície cortical produz ondas positivas em superfície pela despolarização do corpo celular e das células piramidais. O gerador C encontra-se a 900 μm , dando lugar a ondas negativas com a despolarização das células piramidais [28].

Em resumo pode dizer-se então que o sinal dentro do sistema nervoso tem origem em dois mecanismos principais: 1º dentro dos nervos e neurónios por meio de potenciais de acção; 2º entre os neurónios por meio da difusão do neurotransmissor através das sinapses. As sinapses e os potenciais de acção resultam em potenciais diferentes em toda a membrana do neurónio. Estes potenciais e os potenciais de repouso das células gliais (são as células mais frequentes do tecido nervoso, ocupam os espaços entre os neurónios, têm funções de isolamento, defesa, modulação da actividade, suporte e nutrição dos neurónios) são os principais contribuintes do sinal de EEG.

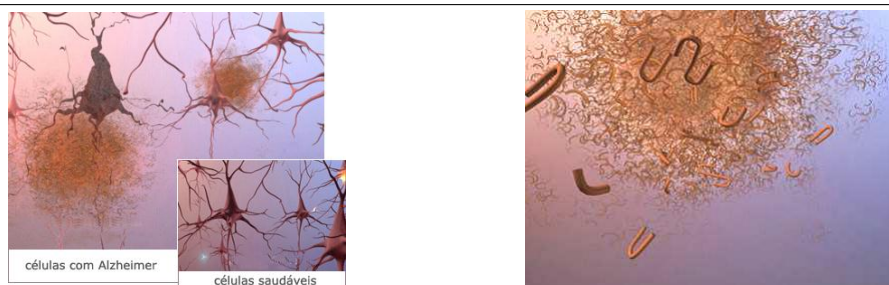
1.6 Doença de Alzheimer

A doença de Alzheimer pode ser entendida como uma demência cortical, degenerativa e progressiva primária do Sistema Nervoso Central, irreversível e de causa desconhecida [29]. Caracteriza-se pela acelerada deterioração intelectual e de comportamento, onde os primeiros sintomas mais comuns são a concentração e os distúrbios de memória sobretudo em relação a acontecimentos mais recentes, subsequentemente outras faculdades mentais e físicas entram em declínio. Nos estágios mais avançados da DA o paciente perde toda a autonomia, assiste-se a uma atrofia geral do cérebro que se caracteriza pela

existência de placas senis e emaranhados neurofibrilares, diminuição da densidade sináptica e morte dos neurónios [30].

No córtex cerebral de pacientes com a DA aparecem placas senis em número excessivo e também se observa a presença de emaranhados neurofibrilares. Outro factor significativo na DA é a presença extremamente reduzida de acetilcolina (ACh), que é necessária para a função cognitiva (é um neurotransmissor essencial para o processamento da memória e aprendizagem). Efectivamente, nesta doença ocorre um menor número de sinapses nas células nervosas, com placas senis formadas pela aglomeração de proteínas anormais entre as células. As placas formam-se quando fragmentos de proteína chamada β -amilóide se aglomeram. A β -amilóide vem de uma proteína maior encontrada na membrana plasmática das células nervosas [30].

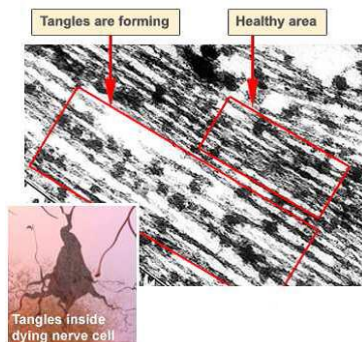
Figura 1.5 Placas de Proteínas β -amiloide.



Fonte:[31]

Os emaranhados neurofibrilares são alterações intracelulares observadas no citoplasma dos neurónios de pacientes com DA. São formados principalmente pela proteína tau, cuja função é estabilizar os microtúbulos dos axónios (estruturas responsáveis pela formação e manutenção dos contactos interneurónios). Estas funções são alteradas quando a proteína tau é modificada pela adição anormal de fósforo no processo. O substrato básico dos emaranhados neurofibrilares é a proteína tau hiperfosforilada (Ver Figura 1.6). Uma anormalidade na proteína tau interrompe a actividade normal das células[32].

Figura 1.6 Emaranhados neurofibrilares.



Fonte:[33]

1.6.1 Factores de risco

Listam-se a seguir os factores de risco:

A Idade: A DA responde por mais de 50% dos casos na faixa etária igual ou superior a 65 anos [34].

O Sexo: Alguns estudos indicam que existe mais casos entre as mulheres. O que pode ter como explicação, o facto de as mulheres viverem mais anos que os homens e não a diferença de sexos [35].

Os Factores genéticos/hereditariedade: Para um número extremamente limitado de famílias, a doença de Alzheimer é uma disfunção genética. Os membros dessas famílias herdaram de um dos pais a parte do DNA (a configuração genética) que provoca a doença [36].

Outros factores: hipertensão; arteriosclerose; tabaco; obesidade; diabetes; doenças cardíacas, traumatismos cranianos, entre outros. Há já muitos dados que sugerem que pessoas com um elevado nível de escolaridade tenham um risco menor do que as que possuem um nível baixo de escolaridade [37].

1.6.2 Sintomas da doença de Alzheimer

A neurodegeneração começa a manifestar-se vinte ou mais anos antes de surgirem os sintomas clínicos. Depois de se manifestarem os sintomas, a duração média da doença é cerca de oito anos. Podem ser consideradas três fases ou estágios da doença (Ver Figura 1.7)

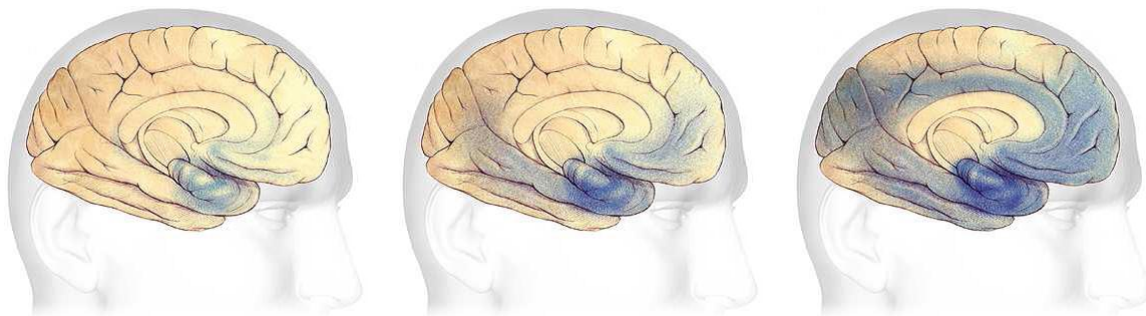
1ª fase inicial: Ocorrem as primeiras mudanças típicas da doença. Os neurónios do lobo temporal e hipocampo são afectados. Registam-se só alguns esquecimentos que não atrapalham a convivência, o doente ainda é independente [38].

2ª fase intermediária leve: São afectadas as zonas do córtex cerebral, o doente começa a depender de outra pessoa, ocorre a perda da memória recente, dificuldade para aprender e reter novas informações, distúrbios de linguagem, dificuldade progressiva para as tarefas diárias, falta de cuidado com a aparência pessoal, irritabilidade, desorientação, apatia, indiferença. Mas não é afectado na locomoção [38]; Intermediária grave - Dificuldade na socialização e perda de memória mais intensa. O paciente é completamente incapaz de aprender e reter novas informações, iniciam-se as dificuldades de locomoção, a comunicação inviabiliza-se e passa a necessitar de cuidados e supervisão integral [38].

3ª fase terminal: produz-se uma atrofia global do cérebro, o doente está completamente dependente de outra pessoa, fica confinado ao seu leito, torna-se num ser vivo vegetativo [39].

Na fase inicial da DA são afectados maioritariamente o lobo temporal, o hipocampo e o córtex cerebral. Com a progressão da DA, os sintomas indicam que os marcadores histológicos, a azul na Figura 1.7, se espalharam pelos lobos frontal e parietal.

Figura 1.7 Estágios: Leve, Moderado e Grave da Doença de Alzheimer.



Fonte:[33]

1.6.3 Diagnóstico

O diagnóstico correcto da DA só pode ser feito por exame do tecido cerebral obtido por biopsia ou necropsia (*post-mortem*) [40]. Como só após a morte do paciente, se pode ter a certeza que este tinha a doença de Alzheimer, o diagnóstico de provável DA é feito excluindo outras causas de demência pelo histórico clínico do paciente (doenças preexistentes, cirurgias, uso de medicações, depressão, perda de memória associada à idade); exames de sangue (hipotireoidismo, deficiência de vitamina b12); Tomografia Computorizada (TC); Tomografia Computadorizada de Perfusão por Emissão de Fóton (SPECT); Tomografia por Emissão de Positrão (PET); Ressonância Magnética Nuclear (RMN) e outros exames. O EEG de rotina e o Electroencefalograma quantitativo (EEGq) têm uso estabelecido como auxiliar na avaliação de demências e encefalopatias. O papel do EEGq como um marcador clínico precoce da doença tem merecido atenção devido aos avanços tecnológicos e à sua relevância prática por ser um método não-invasivo [41]. Para o diagnóstico da doença também são realizados testes padronizados com vista a avaliar a função cognitiva [42].

Existem alguns marcadores, geralmente identificados a partir de exame de sangue, como a apolipoproteína E (APOE), cujos resultados podem mostrar hipótese aumentada de DA. O desenvolvimento de biomarcadores pode melhorar as possibilidades de diagnóstico precoce e poder permitir aos clínicos diagnosticar e tratar a doença de Alzheimer antes de a pessoa apresentar qualquer sintoma [1].

1.6.4 Tratamento

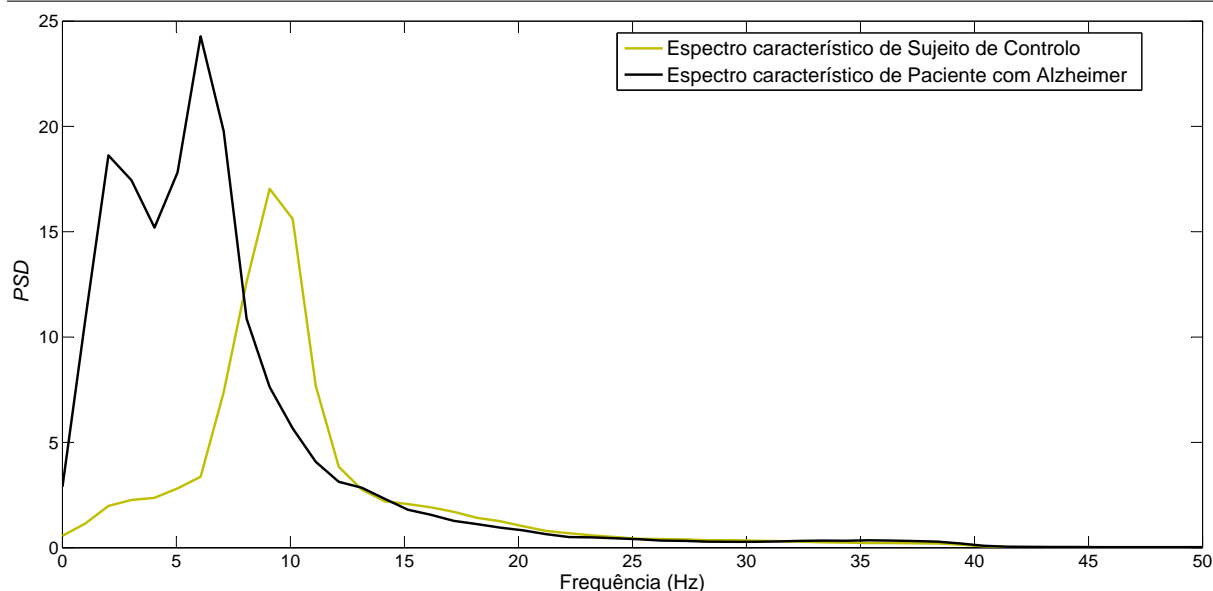
É impossível curar a doença de Alzheimer porque as células nervosas mortas não se podem regenerar. Como tal, apenas se podem tratar os portadores da DA e para isso, utilizam-se geralmente três medicamentos com a mesma acção farmacológica e que apresentam poucos efeitos secundários: a donepesila, a rivastigmina e a galantemina para melhorar a cognição, retardar a evolução, tratar os sintomas e as alterações de comportamento [43].

Num estudo que foi publicado em Setembro de 2010 na revista Neurologia, Robert SW Wilson da Rush University Medical Center, em Chigago, mostra que mentalmente estimulado com actividades como as palavras cruzadas, jogos de tabuleiro e leitura, pode-se deter o declínio de habilidades cognitivas em pessoas com predisposição para a DA. A actividade física pode auxiliar o processo de tratamento da DA, prolonga a qualidade e a duração de uma vida activa, melhorando a coordenação motora e neuromuscular dos portadores da doença [44]. Estudos permitiram concluir que os portadores de DA que ingerem vitamina E em 2000 unidades/dia, aparentemente vivem mais do que aqueles que têm a doença e não fazem uso do suplemento, devido à acção antioxidante dessa vitamina que produz um desaceleramento do processo de envelhecimento celular, retardando o desenvolvimento da DA [45, p. 3].

1.6.5 Diferenças no sinal de EEG de um paciente com DA em relação a um sujeito de controlo

O EEG apresenta bandas de interesse diferentes, tais como delta (1-4 Hz), teta (4-8 Hz), alfa (8-13Hz), beta1 (13-19 Hz), beta2 (19-30 Hz) e gama (30-40 Hz). A DA afecta essas diferentes bandas de frequência. O principal efeito é a “desaceleração” do EEG. Muitos estudos têm mostrado que a DA faz abrandar os sinais de EEG [46]. A DA está associada a um aumento da potência nas bandas de baixa frequência (delta e teta) e a uma diminuição da potência nas frequências mais altas (alfa e beta) [46], conforme mostra a Figura 1.8.

Figura 1.8 Espectro característico de um paciente com Alzheimer e de um sujeito de controlo.



Acredita-se que as primeiras mudanças, após o aparecimento da DA, estão relacionadas com o aumento da actividade em teta e com a diminuição da actividade em beta, seguidos por uma diminuição

da actividade em alfa. Este fenómeno chamado “shift to the left” aparece no estado avançado e intermédio da DA, e pode ser visto notoriamente, no pico máximo de potência que ocorre no domínio de alfa, quando comparado com o grupo de controlo [47]. No entanto, nos estágios iniciais da DA os EEG podem apresentar frequências normais no espectro de potência [48]. Delta aumenta a sua actividade mais tarde, durante o curso da doença. Pacientes com demência grave têm uma diminuição da actividade em alfa e aumento da actividade em delta [46]. A origem fisiopatológica da “desaceleração” do EEG na DA pode ser devida ao défice colinérgico. O cérebro de pacientes com DA mostra uma redução significativa dos marcadores de transmissão colinérgica (perda do neurotransmissor acetilcolina). Vários estudos têm demonstrado que o défice de acetilcolina provoca uma actividade de EEG descoordenada [46].

1.7 Organização da Dissertação

Esta Dissertação vai ser estruturada em seis capítulos:

- Capítulo 1 - Faz a introdução ao tema do Estudo, esclarece os principais objectivos, motivação, dificuldades e soluções. Apresenta o estado da arte, a DA, o EEG e especifica a organização da Dissertação.
- Capítulo 2 - Descreve as técnicas que vão ser utilizadas para analisar os sinais de EEG.
- Capítulo 3 - Apresenta as técnicas de Classificação utilizadas e as técnicas estatísticas de apoio à classificação.
- Capítulo 4 - Descreve os processos de selecção/pré-processamento dos sinais de EEG utilizados neste Trabalho. Expõe as várias metodologias/variantes e respectivos resultados, de vários processos de selecção de características/processamentos espectrais do sinal de EEG.
- Capítulo 5 - Mostra a discussão dos resultados.
- Capítulo 6 - Apresenta as conclusões mais relevantes e as futuras linhas de investigação.

Processamento do Sinal de EEG

Este capítulo apresenta uma visão geral de conceitos teóricos matemáticos que são de grande importância para a compreensão das técnicas de extração de características como o processamento baseado nas transformadas de Fourier e de Wavelet.

As técnicas de processamento digital de sinais têm vindo a ganhar espaço em diversas aplicações, tais como: processamento de voz, áudio, imagem e vídeo, comunicações, automação e controle de processos, robótica, visão computacional, sismologia, meteorologia, finanças, economia, sistemas embutidos, instrumentação, reconhecimento e identificação de padrões, sistemas especialistas, navegação aeroespacial, guerra electrónica, agricultura de precisão, biomecânica, química, biomédica, medicina, biologia e todas as áreas em que é possível utilizar um computador digital para processar informações [49].

No âmbito da engenharia biomédica é visível um grande avanço tecnológico resultante dos microprocessadores digitais, como por exemplo os electroencefalógrafos digitais, os equipamentos de tomografia computadorizada e os scanners de ressonância magnética nuclear.

2.1 Sinal Biomédico

Um sinal é definido como uma variação de uma corrente ou tensão eléctrica ou de outra grandeza utilizado para representar informação. Se for gerado num sistema biológico ou com fins de diagnóstico passa a designar-se sinal biomédico [50, p.52]. O estudo destes sinais proporcionou o aparecimento de diversas áreas da medicina moderna, permitiu uma maior proximidade da medicina com a engenharia biomédica numa tentativa de produzir novos métodos de tratamento e de diagnóstico.

É de salientar que a maior parte dos sinais biológicos é contínua nas suas fontes. Por exemplo a pressão do fluido num vaso sanguíneo ou a diferença de potencial eléctrico através da excitação de membranas do nervo e fibras musculares, são variáveis contínuas que existem em qualquer instante de tempo. Esses sinais são séries temporais de medidas invasivas (ou não), obtidas do corpo humano, representam a actividade dos sistemas fisiológicos e possibilitam a sua caracterização. Se esses sinais

forem extraídos e armazenados de maneira contínua, como acontece num gráfico, a cópia resultante da realidade biológica é conhecida como um sinal contínuo [51].

Existem inúmeros exemplos de sinais biomédicos: electroencefalograma, o electrocardiograma e o electromiograma (EMG) baseados na actividade eléctrica de superfície corpórea. Outros sinais vêm da locomoção, da respiração, do piscar de olhos, da pressão arterial, entre outros. Com computadores cada vez mais sofisticados, capazes de ampliar cada vez mais a capacidade de processamento e armazenamento computacional, o processamento destes sinais por software tem vindo a ganhar força e destaque devido à sua maior aplicabilidade e facilidade de implementação. Neste trabalho utilizam-se registos electromagnéticos que são sinais biomédicos não estacionários [52]. Isto implica que as suas características espectrais dependem do instante de tempo em que se observa o sinal. O motivo da não estacionariedade do sinal deriva do pressuposto que todos os registos da actividade biológica têm sobreposto ruído, o qual pode provir de diversas fontes (equipamentos de registo, sistemas interferentes externos ou mesmo a própria actividade biológica) [53].

2.1.1 Ruídos possíveis em Sinais Biomédicos

2.1.1.1 Ruídos devido aos instrumentos utilizados

Devido aos baixos níveis de amplitude da maioria dos sinais biomédicos, podem ser necessários altos factores de amplificação (na ordem de algumas centenas ou milhares de vezes). Assim, os ruídos electrónicos dos amplificadores de instrumentação também são amplificados. A redução destes ruídos pode ser feita através da refrigeração dos dispositivos a temperaturas muito baixas (não é prático e é caro), da utilização de fontes e energia de baixo ruído, e por último, da utilização de amplificadores electrónicos com alta impedância de entrada e alta taxa de rejeição em modo comum [54].

2.1.1.2 Ruídos devido ao ambiente de experimentação

As ondas electromagnéticas como as ondas de rádio, da TV, radiação de lâmpadas fluorescentes e de monitores são captadas pelos cabos, conectores e demais dispositivos do sistema, levando assim ao aparecimento de ruído no sinal em frequências de 50 Hz ou de 60 Hz, dependendo das fontes de energia. As interferências ambientais podem ser atenuadas através da utilização de cabos com blindagem electromagnética [54].

2.2 Transformada de Fourier

Uma forma de analisar o espectro do sinal EEG é estudar as componentes em frequência que o formam. Neste sentido, a Transformada de Fourier (FT) utiliza-se para extrair informação espectral de

sinais estacionários. Os coeficientes desta transformada são o resultado da computação de uma série de produtos internos do sinal com uma série de funções sinusoidais de duração infinita. No caso de um sinal de EEG $x(t)$ será amostrado a uma frequência de amostragem f_s em que f_s é o inverso do período de amostragem, dando origem a um sinal discreto: $x[n] = x(n \cdot Ts)$, com $n = 0, \dots, N-1$, de N amostras. Da aplicação da transformada discreta de Fourier (DTFT, *Discrete-Time Fourier Transform*) resulta o espectro do sinal $X[k], k = 0, \dots, N-1$. Desta forma assume-se que o sinal de partida $x[n]$ é periódico, com um período fundamental N , e frequência fundamental $\Omega = 2\pi/N$. A variável k é dada pela frequência do sinusóide associado ao $X[k]$, sendo então a frequência discreta $\Omega = k\Omega_0$. As seguintes fórmulas ilustram a análise por DTFT, associada ao sinal [55].

$$x[n] = \frac{1}{N} \cdot \sum_{k=0}^{N-1} X[k] \cdot e^{j \cdot k \cdot \Omega_0 \cdot n}, n = 0, \dots, N-1 \quad (2.1)$$

$$X[k] = \frac{1}{N} \cdot \sum_{n=0}^{N-1} x[n] \cdot e^{-j \cdot k \cdot \Omega_0 \cdot n}, k = 0, \dots, N-1 \quad (2.2)$$

Segundo as equações anteriores, os coeficientes do sinal $x[n]$ são dados pelos $X[k]$. Ambos constituem um par DTFT, pelo que a partir dos N valores de $X[k]$ é possível recuperar $x[n]$, e vice-versa. Tanto $x[n]$ como $X[k]$ proporcionam uma descrição completa do sinal. Como já foi mencionado a representação de Fourier é a única que se pode implementar numericamente num computador e só se torna possível aproximar as outras representações mediante o processamento por DTFT. Uma forma de analisar as características de um sinal consiste em observar a distribuição da potência no domínio da frequência. Para isso, utiliza-se a função Densidade Espectral de Potência (*PSD - Power Spectral Density*), em que no cálculo se considera que o sinal analisado é de potência [56]. Uma maneira comum de estimar a *PSD* é calculá-la como a FT da função de autocorrelação do sinal EEG, uma vez que ambos são um par de transformada de Fourier [56]. Neste sentido, a seguinte equação ilustra uma estimativa da função de autocorrelação $R_{xx}[u]$, a partir do sinal discreto de $x[n]$:

$$R_{xx}[u] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-m-1} x[n] \cdot x[n+u] & \text{se } u \geq 0 \\ R_{xx}^*[-u] & \text{se } u < 0 \end{cases} \quad (2.3)$$

A equação a seguir representa a estimativa da *PSD*, $S_x[k]$, onde $DTFT \{x\}$ representa a Transformada Discreta de Fourier e Ω a frequência discreta.

$$\begin{aligned} S_x[k] &= \frac{1}{N} \cdot DTFT \{R_{xx}[u]\} = \frac{1}{N} \sum_{u=0}^{2N-1} R_{xx}[u] \cdot e^{-j \cdot \Omega \cdot u} \\ &= \frac{1}{N} \sum_{u=0}^{2N-1} R_{xx}[u] \cdot e^{-j \cdot \frac{2\pi k}{2N-1} \cdot u}, k = 0, \dots, 2 \cdot N - 1 \end{aligned} \quad (2.4)$$

Apesar da estimativa da *PSD* se poder calcular mediante outros algoritmos, este permite uma resolução espectral maior em relação à obtida com outras técnicas não paramétricas.

2.3 Transformada de Fourier de Curta Duração

Alguns problemas surgem quando se utiliza a FT, pois esta não é capaz de fornecer informações das características do sinal simultaneamente no tempo e na frequência, quando o sinal varia ao longo do tempo (não estacionário). As componentes espectrais não reflectem as mudanças ao longo do tempo, o que é um problema na análise de sinais não-estacionários [52, 57]. Para resolver este problema foi introduzida a Transformada de Fourier de Curta Duração (STFT - *Short-Time Fourier Transform*). A STFT usa uma janela de comprimento constante que desliza pelo sinal e transforma o sinal original contido no espaço temporal desta no domínio da frequência. A janela é então transferida para uma nova posição e esta transformação é repetida numa parte diferente do sinal [52, 57].

A equação de análise da STFT assenta na suposição de que o segmento de sinal recolhido por cada janela é estacionário, o que depende em grande parte da janela utilizada, $w[n]$, [58]. A seguinte fórmula representa a equação de análise da STFT discreta, onde L é o tamanho da janela, N o número de amostras, $N_T = \frac{N}{L}$ e $w[n]$ a função da janela não nula no intervalo $[0, L - 1]$.

$$X_w[m, k] = \sum_{n=0}^{N-1} x[n] \cdot w^*[n - m] \cdot e^{-j \cdot k \cdot \Omega_0 \cdot n}, k = 0, \dots, L - 1; m = 0, \dots, N_T - 1 \quad (2.5)$$

Novamente tal como ocorreu com a TF uma forma de analisar as propriedades de um sinal é estudar como a potência é distribuída no plano tempo-frequência. Assim como no caso do FT, são estimadas inicialmente a função de autocorrelação de curto período de tempo (STA, *Short-Time Autocorrelation*) para sequências finitas, onde $w'[n] = w[n]$.

$$R_{xx}[u] = \begin{cases} \frac{1}{L} \sum_{n=0}^{L-u-1} x[n + m] \cdot w'[n] \cdot x[n + m + u] \cdot w'[u + m] & \text{se } u \geq 0 \\ R_{xx,w}^*[-u] & \text{se } u < 0 \end{cases} \quad (2.6)$$

De seguida, calcula-se a *PSD*, $S_{x,w}[m, k]$, como uma transformada de Fourier no domínio tempo-frequência, que consiste na aplicação da DFT à STA.

$$\begin{aligned} S_{m,k}[k] &= \frac{1}{L} \cdot DTFT \{R_{xx,w}[u]\} = \frac{1}{L} \sum_{u=0}^{2L-1} R_{xx,w}[u] \cdot e^{-j \cdot \Omega \cdot u} \\ &= \frac{1}{L} \sum_{u=0}^{2L-1} R_{xx,w}[u] \cdot e^{-j \cdot \frac{2\pi k}{2L-1} \cdot u}, k = 0, \dots, 2 \cdot L - 1; m = 0, \dots, N_T - 1 \end{aligned} \quad (2.7)$$

2.4 Transformada Wavelet

Os sinais biomédicos, em geral, tendem a ter oscilações rápidas, simultaneamente, em curtos intervalos de tempo ou variações lentas em intervalos longos [57]. Como resultado, é conveniente usar uma transformação que oferece uma resolução variável no plano tempo-frequência. Este é o caso da transformada Wavelet (WT, *Wavelet Transform*), que oferece uma boa resolução em tempo para as altas frequências e boa resolução em frequência para baixas frequências [57]. Na seguinte equação consegue-se discernir o processo, onde τ é o parâmetro de translação e a representa a escala das funções base.

$$\Psi_{\tau,a} = \frac{1}{\sqrt{|a|}} \cdot \Psi\left(\frac{t-\tau}{a}\right); \quad (2.8)$$

A escala (a), actua como um factor de ampliação ($a>1$) ou de compressão ($a<1$), dependendo se a análise inclui os comportamentos que se estendem no tempo ou os detalhes específicos, respectivamente [57]. Estas ideias sugerem que é possível estabelecer uma relação entre o conceito de escala e frequência. Assim, as altas escalas correspondem às frequências baixas, apanhando as variações que afectam todo o sinal. Por outro lado, as baixas escalas estão relacionadas com as componentes de alta frequência, reflectindo as flutuações de curto duração do sinal [57]. Como os sinais iniciais são discretos, de seguida, passa-se a detalhar a variante da WT mais amplamente utilizada para este tipo de análise. Entre as várias representações, a DWT é a mais utilizada. Isso deve-se à sua capacidade de se adaptar a diferentes problemas, e ao seu baixo custo computacional. A DWT usa uma análise multiresolução que consiste na obtenção de versões do sinal com resolução cada vez menor através de sucessivas filtragens. Para isso, a DWT usa dois conjuntos de funções: a função de escala, $\varphi[n]$, e a função Wavelet, $\psi[n]$. Cada uma está associada a um filtro passa-baixo, $h[n]$, e passa-alto $g[n]$, respectivamente [57].

$$\varphi[n] = \sum_k h[k] \cdot \varphi[2 \cdot n - k] \quad (2.9)$$

$$\psi[n] = \sum_k g[k] \cdot \varphi[2 \cdot n - k] \quad (2.10)$$

Além disso, uma função de escalonamento, $\varphi_{(j,k)}[n]$ ou uma função Wavelet, $\psi_{(j,k)}[n]$, discretizadas com um parâmetro de escala j e um de translação k , são obtidos a partir de $\varphi_{0,0}[n] = \varphi[n]$ e $\psi_{0,0}[n] = \psi[n]$ [57],

$$\varphi_{j,k}[n] = 2^{-\frac{j}{2}} \cdot \varphi[2^{-j} \cdot n - k], \text{ com } \varphi_{0,0}[n] = \varphi[n] \quad (2.11)$$

$$\psi_{j,k}[n] = 2^{-\frac{j}{2}} \cdot \psi[2^{-j} \cdot n - k], \text{ com } \psi_{0,0}[n] = \psi[n] \quad (2.12)$$

Diferentes escalas e traduções destas funções permitem obter diferentes localizações em tempo e frequência do sinal. Os sinais $h[n]$ e $g[n]$ são as respostas impulsivas dos filtros passa-baixo e passa-alto utilizados na análise wavelet. Elas definem o tipo de wavelet usada na análise. Por isso, a decomposição do sinal em diferentes bandas de frequência é conseguida por sucessivos filtros passa-baixo e passa-alto do sinal filtrado no tempo, seguido por uma subamostragem por um factor de dois [57]. O processo de filtragem, seguido de uma subamostragem, até ao nível máximo de decomposição ($\log_2(N_T)$) é ilustrado nas seguintes equações, onde $DWT_{aprox}[0, n] = x[n]$:

$$DWT_{aprox}^{\phi}[j, k] = \sum_{n=0}^{\frac{N}{2^{j-1}}} \cdot DWT_{aprox}^{\phi}[j-1, n] \cdot h[2 \cdot k - n] \quad (2.13)$$

$$DWT_{det}^{\psi}[j, k] = \sum_{n=0}^{\frac{N}{2^{j-1}}} \cdot DWT_{aprox}^{\phi}[j-1, n] \cdot g[2 \cdot k - n] \quad (2.14)$$

$$sendo : k = 0, \dots, \frac{N}{2^j}; j = 0, \dots, \log_2(N)$$

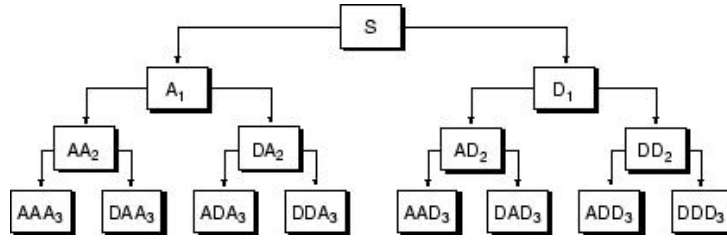
onde, $DWT_{aprox}^{\phi}[j, k]$ e $DWT_{det}^{\psi}[j, k]$ são as saídas do filtro passa-alto e passa-baixo, respectivamente após a sub-amostragem. $DWT_{det}^{\psi}[j, k]$, representa os coeficientes de detalhe nível j , enquanto que $DWT_{aprox}^{\phi}[j, k]$ são os coeficientes de aproximação do nível j . Este processo, chamado de codificação sub-banda, pode ser repetido n vezes, até que não seja possível diminuir mais a resolução do sinal. Em cada nível, obtém-se versões do sinal original com metade da resolução temporal (devido à subamostragem) com o dobro da escala (devido à filtragem), o que permite que o sinal possa ser analisado em diferentes faixas de frequência com resoluções diferentes [57].

Finalmente, o sinal original pode ser reconstruído a partir dos coeficientes de aproximação do último nível de decomposição ($J = \log_2(N)$) e dos coeficientes de detalhes de todos os níveis ($j = 1, \dots, J$). Essa ideia é ilustrada pela seguinte equação:

$$x[n] = \sum_{k=0}^{\frac{N}{2^J}} DWT_{aprox}^{\phi}[J, k] \cdot \phi_{J,k}[n] + \sum_{j=1}^J \sum_{k=0}^{\frac{N}{2^j}} DWT_{det}^{\psi}[j, k] \cdot \psi_{j,k}[n] \quad (2.15)$$

De forma a obter a PSD pela WT calculou-se o Espectro de Wavelet a partir dos coeficientes da Wavelet Packet Transform (WPT) correspondentes aos nós terminais. Ao contrário da DWT, a WPT permite uma resolução não só nas baixas frequências mas também nas altas frequências pois a sua árvore de decomposição é diferente da árvore da DWT (detalhes são divididos em coeficientes de aproximação e de detalhe como mostra a Figura 2.1).

Figura 2.1 Exemplo de Árvore de decomposição da WPT até ao nível 3.



Fonte:[59]

Primeiro de tudo, a WPT, é definida como [60],

$$X_{\varphi,\psi}[j,m,k] = \sum_{n=0}^{\frac{N}{2^{j-1}}} WP^{\varphi,\psi}[j,m,n] \cdot w_{j,m,k}[n],$$

$$w_{j,m,k}[n] = 2^{-\frac{j}{2}} \cdot [2^{-j} \cdot n - k], j = 1, \dots, J; m = 0, 1, \dots, 2^j - 1; k = 1, \dots, 2^{J-j} \quad (2.16)$$

onde, $J = \log_2(N_T)$, j representa o nível da WPT. Agora está-se na condição de definir o espectro de WPT pelo periodograma da WPT do sinal baseado nas observação de tamanho $N = 2^j$.

$$S_{x,\varphi,\psi}[m,k] = \left[\sum_{j=1}^J X_{\varphi,\psi}[j,m,k] \right]^2 \quad (2.17)$$

De modo a obter o intervalo de frequência do espectro de Wavelet em cada um dos níveis de decomposição usa-se o intervalo $[F_s/2^{(j+1)}, F_s/2^{(j)}], j = 1, \dots, \log_2(N)$. De seguida, os valores absolutos dos coeficientes de Wavelet foram calculados e ordenados por frequência.

Tabela 2.1 Propriedades das Famílias mais importantes de Wavelet

Propriedades	Famílias de Wavelet						
	Db	Sym	Coif	Bior	Morl	Meyr	Haar
Regularidade infinita					x	x	
Regularidade arbitrária	x	x	x	x			
Simétrica				x		x	x
Número Arbitrário de momentos zero	x	x	x	x			
Assimétrica	x						
Perto de ser simétrica		x	x				
Análise Ortogonal	x	x	x	x		x	x
Análise Biortogonal	x	x	x	x		x	x
Reconstrução exacta	x	x	x	x		x	x
Filtros FIR	x	x	x	x			x
Transformada contínua	x	x	x	x	x	x	x
Transformada discreta	x	x	x	x		x	x
Algoritmo rápido	x	x	x	x			

Db=Daubechies;Sym=Symlet;Coif=Coiflet;Bior=Biortogonal;Morl=Morlet;Meyr=Meyer;Haar=Haar;

Procedeu-se a uma pré selecção das Famílias de Wavelets utilizadas neste trabalho, que serão a Wavelet Biortogonal, a Wavelet Daubechies, a Wavelet Symlet e a Wavelet Coiflet, porque à partida serão aquelas que nos permitirão processar o sinal mais rapidamente, diminuindo o tempo de processamento do sinal e ao mesmo tempo apresentam as características necessárias para se ajustarem com eficácia ao Sinal de EEG, como: Análise biortogonal, análise ortogonal, reconstrução exacta, número arbitrário de momentos zero entre outras (Consultar Tabela 2.1).

2.5 Normalização da Densidade Espectral de Potência

De forma a poder simplificar o processo de cálculo de algumas características procedeu-se à normalização da PSD entre 0 e 1.

$$PSD^{(i)}[m, k] \begin{cases} \frac{S_{x,w}[n]}{\sum_{n=k_1}^{k_2} S_{x,w}[n]} & m = 0, \dots, N_T - 1 \\ \frac{S_{x,\varphi,\psi}[n]}{\sum_{n=k_1}^{k_2} S_{x,\varphi,\psi}[n]} & m = 0, \dots, 2^j - 1 \end{cases} \quad (2.18)$$

onde $i = \{STFT, WT\}$, k_1 e k_2 representam frequências discretas de corte e eles podem ser calculadas como:

$$k_1 = \frac{(f_1 \cdot N_T)}{F_s}; k_2 = \frac{(f_2 \cdot N_T)}{F_s} \quad (2.19)$$

onde f_1 e f_2 são as frequências contínuas de corte de um filtro digital. F_s é a frequência de amostragem e finalmente N_T representa as amostras do sinal [61]. Como a $PSD^{(i)}[m, k]$ neste trabalho é calculada primeiro para um conjunto de segmentos e geralmente em seguida mediada para os eléctrodos e para os indivíduos ela assume a seguinte forma,

$$\langle PSD_n^{(i)}[k] \rangle = \begin{cases} \frac{1}{N_T} \sum_{m=0}^{N_T-1} PSD_n^{i=STFT}[k] \\ \frac{1}{N_T} \sum_{m=0}^{2^j-1} PSD_n^{i=WT}[k] \end{cases} \quad (2.20)$$

onde N_T representa o número de amostras. De forma a simplificar o processo relacionou-se k com f através das equações anteriores da seguinte forma.

$$\langle PSD_n^{(i)}(f) \rangle = \langle PSD_n^{(i)} \left[\frac{(k \cdot F_s)}{N_T} \right] \rangle \quad (2.21)$$

2.6 Selecção de características espectrais

A fim de realizar uma análise aprofundada das características espectrais do EEG dos registos dos pacientes diagnosticados com doença provável de Alzheimer e dos sujeitos de controlo, neste trabalho os seguintes parâmetros são calculados utilizando o processamento, das transformadas de STFT e de WT,

descritas anteriormente. A frequência média (*MF*) oferece um meio simples de resumir todo o espectro do conteúdo da *PSD*. É definida como a frequência que compreende 50% da potência.

$$0.5 \sum_{1Hz}^{40Hz} PSD_n(f) = \sum_{1Hz}^{MF} PSD_n(f) \quad (2.22)$$

Nos indivíduos normais as oscilações na banda alfa são dominantes. Então, a *PSD* apresenta um pico máximo em torno da banda alfa (8-13 Hz). Neste momento está-se na condição de calcular um outro parâmetro denominado Frequência Individual de Alfa (*IAF - Individual Alpha Frequency*). Para detectar a *IAF*, foi calculada a frequência média na faixa alfa estendida (4-15 Hz), pois observou-se no espectro da *PSD* que existiam vários picos (máximos) nesta faixa.

$$0.5 \sum_{4Hz}^{15Hz} PSD_n(f) = \sum_{4Hz}^{IAF} PSD_n(f) \quad (2.23)$$

Semelhante à *MF*, a frequência de fronteira espectral 95% (*SEF95% - Spectral Edge Frequency 95%*) é estimada como a frequência que compreende 95% da potência do espectro. Assim, utilizou-se a *SEF95%* como um indicador do limite superior no conteúdo espectral da *PSD_n*. A equação seguinte representa a expressão a partir da qual a *SEF95%* é calculada.

$$0.95 \sum_{1Hz}^{40Hz} PSD_n(f) = \sum_{1Hz}^{SEF95\%} PSD_n(f) \quad (2.24)$$

Finalmente, a entropia espectral (*SE - Spectral Entropy*) foi estimada para medir o nivelamento da *PSD*. Ela representa uma medida de desordem, que pode ser usada como uma estimativa de irregularidade do EEG [62]. Para calcular a *SE*, foi aplicada a definição de entropia de Shannon calculada sobre a função da *PSD_n*.

$$SE = - \sum_{1Hz}^{40Hz} PSD_n(f) \cdot \text{Log}[PSD_n(f)] \quad (2.25)$$

A Potência Relativa (*RP - Relative Power*) calcula-se através do somatório das componentes da *PSD_n* nas bandas de frequências clássicas teta(1-4 Hz), delta(4-8 Hz), alfa(8-13Hz), beta1(13-19 Hz), beta2(19-30 Hz) e gama(30-40 Hz).

$$RP(delta) = \sum_{1Hz}^{4Hz} PSD_n(f) \quad (2.26)$$

$$RP(teta) = \sum_{4Hz}^{8Hz} PSD_n(f) \quad (2.27)$$

$$RP(alfa) = \sum_{8Hz}^{13Hz} PSD_n(f) \quad (2.28)$$

$$RP(beta1) = \sum_{13Hz}^{19Hz} PSD_n(f) \quad (2.29)$$

$$RP(beta2) = \sum_{19Hz}^{30Hz} PSD_n(f) \quad (2.30)$$

$$RP(gamma) = \sum_{30Hz}^{40Hz} PSD_n(f) \quad (2.31)$$

O equilíbrio de potências entre as altas e as baixas frequências foi calculado para resumir a desaceleração do espectro dos pacientes de Alzheimer do EEG. Para enfatizar as diferenças locais do espectro combina-se a *RP* das bandas de frequência convencionais [63].

$$r_1 = \frac{RP(alfa)}{RP(teta)} \quad (2.32)$$

$$r_2 = \frac{RP(alfa) + RP(beta1) + RP(beta2) + RP(gama)}{RP(delta) + RP(teta)} \quad (2.33)$$

$$r_3 = \frac{RP(beta1) + RP(beta2)}{RP(teta)} \quad (2.34)$$

$$r_4 = \frac{RP(beta2)}{RP(delta)} \quad (2.35)$$

Assim, a r_1 é capaz de detectar mudanças nos estágios iniciais da doença, quando um ligeiro abrandamento no espectro do EEG aparece. r_2 foi usado para obter um índice que resume a desaceleração global do EEG. Finalmente, algumas obras literárias/artigos utilizam a relação $RP(alfa)/RP(delta)$ para detectar alterações em estágios avançados da doença de Alzheimer. Neste estudo, substitui-se na formula em r_1 a potência de alfa por potência de beta constituído duas novas relações r_3 e r_4 . Esta modificação é apoiada pelo trabalho [63], onde as mudanças mais significativas na actividade de EEG espontâneo entre pacientes com DA e os EEG de sujeitos de controlo foram observadas em $RP(beta2)$. Por isso, não se definiu apenas um índice para sintetizar a informação de toda a banda beta (r_3), mas também uma medida mais específica, que quantifica a contribuição da banda beta2 (r_4) [63].

Métodos de Classificação

Este capítulo estrutura-se em três partes fundamentais. A primeira parte apresenta as características básicas inerentes às RNAs: o tipo de Arquitectura de RNA; os métodos de aprendizagem das RNAs; o conceito de função de activação de uma RNA; os Métodos de Treino das RNAs Supervisionadas; uma abordagem à aplicação das RNAs em diferentes áreas de conhecimento, dando especial destaque à Biomédica; um registo cronológico do desenvolvimento das RNAs. A segunda parte descreve o conceito de Regressão Logística. Finalmente, a terceira parte, enuncia alguns métodos de apoio à decisão: Curvas ROC (*Receiver Operating Characteristic*) e os testes não paramétricos de Kruskal-Wallis, de Levene e de Kolmogorov-Smirnov(KS).

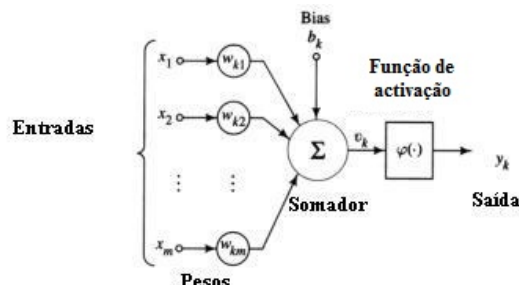
3.1 Redes Neurais Artificiais

O estudo das Redes Neurais Artificiais tem sido relevante para os investigadores de inteligência artificial. As RNAs têm vindo a constituir cada vez mais, métodos válidos para a solução de problemas complexos de difícil resolução, através de inteligência artificial, assumindo-se como uma valiosa ferramenta para diversas aplicações. As RNAs são dispositivos que permitem tolerância a ruídos e falhas, o que advém da sua capacidade de generalização, sendo capazes de suportar ruídos e imperfeições no ambiente, e ainda assim, mantêm um resultado desejado [64]. As RNAs são por definição sistemas computacionais estruturados numa aproximação à computação baseada em ligações. Nós simples (neurónios, processadores ou unidades) são interligados para formar uma rede de nós - daí o termo rede neuronal. No interior dos neurónios artificiais, o valor de offset (bias (b_k)) e as entradas ponderadas pelos seus respectivos pesos sinápticos (w_{km}) são somados. O valor resultante é então submetido a uma função de activação que fornece a saída do neurónio k , onde x_j é a j -ésima entrada do neurónio; w_{kj} é o peso associado a esta conexão; b_k é o bias, desvio ou offset e φ é a função de activação (ver Figura 3.1).

O conjunto das conexões existentes entre os neurónios de uma rede neuronal forma o vector de pesos, cuja dimensão é definida pela arquitectura da rede. O modo segundo o qual os neurónios de uma rede

neuronal artificial estão estruturados está intimamente ligado com o algoritmo de aprendizagem usado no treino da rede neuronal artificial. Os neurónios estão organizados graficamente e ligados entre si com uma orientação, cuja estrutura define a arquitectura adoptada.

Figura 3.1 Neurónio Artificial.



Fonte: [65]

A saída do nó, y_k , é dada pela seguinte expressão:

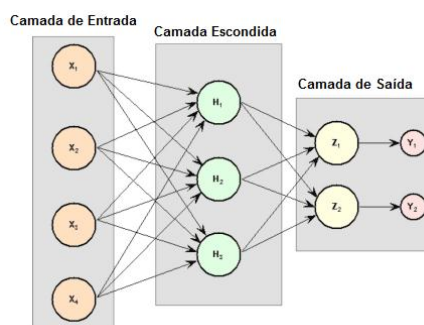
$$y_k = \varphi \left(\sum_{i=1}^m w_{ki}x_i + b_k \right) \quad (3.1)$$

3.1.1 Tipo de arquitecturas para Redes Neurais Artificiais mais utilizadas

3.1.1.1 Redes Feed-forward

São Redes com uma ou mais camadas ligadas em sequência, a propagação de sinais nesta rede é puramente unidireccional, isto é, os sinais são propagados apenas da entrada para a saída, e nunca vice-versa. A arquitectura Feed-forward é caracterizada pela organização dos neurónios em camadas separadas hierarquicamente: camada de entrada, camadas escondidas e camada de saída (Figura 3.2).

Figura 3.2 Rede Neuronal Feed-forward.



Baseado em [66]

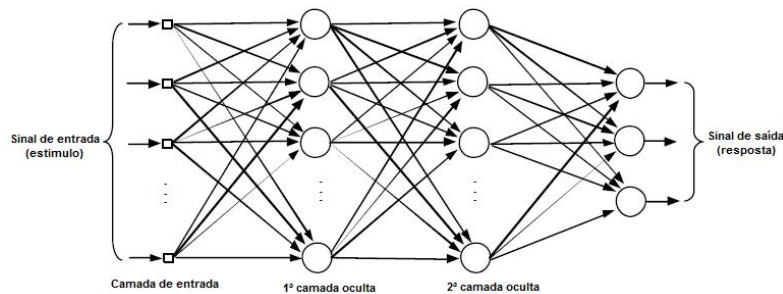
No caso mais simples de redes em camadas (Layers) tem-se uma camada de entrada com neurónios cujas saídas alimentam a última camada da rede. Podem ser treinadas por um algoritmo de aprendizagem

simples (geralmente regra-delta). Este algoritmo calcula os erros entre a saída dos dados calculados e a saída desejada de forma a ajustar os pesos (Confrontar Figura 3.2).

3.1.1.1 Redes Feed-forward com multicamadas

Esta classe de rede consiste em múltiplas camadas de unidades computacionais (neurónios) interligadas (ver Figura 3.3). Isso quer dizer que cada neurónio tem uma camada com ligações directas a neurónios da próxima camada. São redes que apresentam uma ou mais camadas intermediárias ou escondidas (camadas em que os neurónios são efectivamente unidades processadoras, mas não correspondem à camada de saída). Adicionando-se uma ou mais camadas intermédias, aumenta-se o poder computacional de processamento não-linear e armazenagem da rede.

Figura 3.3 Redes Neurais tipo Feed-Forward com múltiplas camadas.



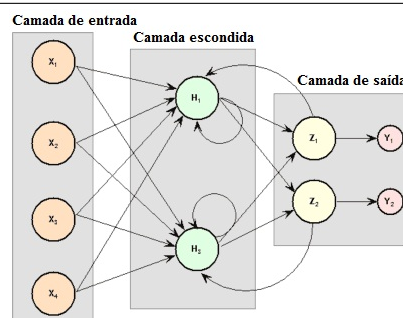
Baseado em Fonte:[67]

O conjunto de saídas dos neurónios de cada camada da rede é utilizada como entrada, para a camada seguinte da mesma. As redes multicamadas usam a técnica de aprendizagem Back-Propagation, que será explicada mais á frente neste trabalho.

3.1.2 Redes Neurais Recorrentes

Redes com ligações entre elementos da mesma camada e/ou com elementos de camadas anteriores (Confrontar Figura 3.4).

Figura 3.4 Rede Neuronal Recorrente.



Fonte: Baseado em [66]

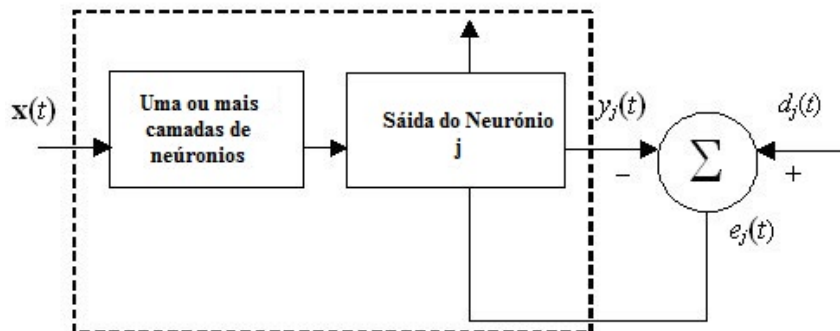
3.1.3 Métodos de aprendizagem das redes neuronais artificiais

Considera-se que uma das propriedades mais importantes das redes neuronais artificiais é a capacidade de aprender com base no seu ambiente de modo a melhorar a sua performance. A aprendizagem realiza-se através de um processo interactivo de ajustes aplicado aos seus pesos, o treino.

3.1.3.1 Aprendizagem supervisionada

Neste tipo de aprendizagem a rede é treinada através da introdução de valores de entrada/saída; para cada entrada a rede produz uma resposta na saída ($y_j(t)$), que é comparada com a resposta desejada ($d_j(t)$) (Ver Figura 3.5). Através da análise de erros ($e_j(t)$), realiza-se o ajuste dos pesos (bias e w) [64].

Figura 3.5 Diagrama do processo de aprendizagem supervisionada.



Baseado na Fonte: [68]

3.1.3.1.1 Processamento Neuronal numa aprendizagem supervisionada

Este processamento pode ser dividido em duas fases, a fase do Recall que consiste no processo de cálculo da saída da rede, dado um certo padrão de entrada - Recuperação da Informação e na fase do Learning que é o Processo de actualização dos pesos sinápticos para a aquisição do conhecimento - Aquisição da Informação.

3.1.3.2 Aprendizagem não supervisionada

Não existe saída desejada. A rede é treinada através de excitações ou padrões de entrada para organizar os padrões em categorias. Para uma entrada aplicada à rede, será facultada uma resposta indicando a classe à qual a entrada pertence; se o padrão de entrada não corresponde às classes existentes é então criada uma nova classe [64].

3.1.4 Função de activação de uma Rede Neuronal Artificial

Uma função de activação executa uma operação matemática no sinal de saída. As funções de activação mais complexas são utilizadas dependendo do tipo de problema em causa (consultar Tabela 3.1)

Tabela 3.1 Funções de activação.

Função	Equação (com polarização)	Gráfico (sem polarização)	Gráfico (com polarização)
Degrau	$y = \begin{cases} 1, & x > -b \\ 0, & x < -b \end{cases}$		
Degrau Simétrico	$y = \begin{cases} 1, & x > -b \\ -1, & x < -b \end{cases}$		
Linear	$y = x + b$		
Logística Sigmoidal	$y = \frac{1}{1 + e^{-(x+b)}}$		
Tangente Sigmoidal	$y = \frac{e^{(x+b)} - e^{-(x+b)}}{e^{(x+b)} + e^{-(x+b)}}$		

3.1.5 Métodos de Treino das redes neuronais artificiais Supervisionadas

O processo de aprendizagem em RNA é um grande problema que envolve a adaptação dos pesos da Rede. Assim, os métodos de optimização numérica têm sido aplicados como uma abordagem para resolver o problema correctamente.

3.1.5.1 Algoritmo Back-Propagation

O algoritmo Back-propagation é um algoritmo de aprendizagem supervisionada baseado na correcção de erros. Foi publicado por Rumelhart, Hinton, Williams e McClelland em 1986 no livro "Parallel Distributed Processing", embora tenha sido proposto por investigadores como Bryson e Ho em 1969,

Werbos em 1974, Parker em 1985 e Minsky e Papert em 1988 com diferentes objectivos. O objectivo do algoritmo Back-propagation é encontrar na superfície de erro um mínimo global, ou seja, valores para os pesos sinápticos que minimizam o erro da rede [64].

O algoritmo de aprendizagem supervisionada Back-propagation utiliza os pares de entrada-saída para ajustar os pesos da rede. A frequência com que estes ajustes são feitos depende do tipo de treino, que pode ser em modo incremental (in-line) onde a actualização é realizada após cada exemplo apresentado; por ciclo (batch ou lote) onde a actualização é realizada após todos os exemplos do conjunto de treino serem apresentados à RNA [64]. Os critérios de paragem, embora ainda não existam critérios com sucesso para terminar o processo de aprendizagem, podem ser: quando, percorrendo-se a superfície de erro, for encontrado um mínimo; quando a variação do erro for infinitamente pequena sendo insuficiente para melhorar a performance da rede; por a rede ter atingido o erro alvo; por o treino ter atingido o número de máximo de iterações e por a rede não aprender mais, mas apesar disso o erro diminui no conjunto de treino (“overfitting”). Usa-se um conjunto de validação (também representativo do problema) ou o parâmetro Weight Decay (definido mais à frente) para parar o treino quando o erro nesse conjunto deixar de diminuir. É comum observar-se na prática a combinação dos critérios acima descritos [56, 64, 69].

O algoritmo Back-propagation tem duas fases: a Forward e a Backward. Na fase Forward, a entrada é apresentada à primeira camada da RNA (camada de entrada C_0). Após os nós da camada $C_i (i > 0)$ calcularem os sinais de saída, estes servem como entrada para a definição das saídas produzidas pelos nós da camada C_{i+1} , e assim consecutivamente. As saídas produzidas pelos nós da última camada são comparadas com as saídas desejadas. Na fase Backward, os nós da camada actual ajustam os seus pesos de forma a reduzir o erro. O erro de um nó das camadas intermediárias é calculado utilizando os erros dos nós da camada seguinte ligados a elas, ponderada pelos pesos das conexões entre elas. Esta fase dá-se a partir da última camada, até chegar à camada de entrada.

3.1.5.2 Algoritmo do Gradiente Conjugado - *Scaled Conjugate Gradient Algorithm*

O algoritmo de gradiente escalado conjugado (SCG) é apontado como uma ferramenta adequada para lidar com problemas de grande escala. O SCG usa uma abordagem de Levenberg-Marquardt, a fim de dimensionar o tamanho do passo, que permite reduzir consideravelmente a complexidade. Portanto, o vector de pesos é actualizado a cada iteração, como mostra a expressão seguinte. [70].

$$\Delta \bar{w} = \alpha_k \bar{p}_k = \frac{-\bar{p}_k E'(\bar{w}_k)}{-\bar{p}_k E''(\bar{w}_k) \bar{p}_k} \quad (3.2)$$

onde, p_k se refere à direcção do conjugado para a descida, α_k é o tamanho do passo para o algoritmo nessa direcção, E' é o vector obtido a partir da primeira derivada do erro (E) em relação a cada componente do vector de pesos e E'' é a matriz obtida a partir da segunda derivada do erro E em relação ao vector

de pesos, ou seja, a matriz Hessiana para o erro (E). O princípio do algoritmo SCG, como ocorre no algoritmo de Levenberg-Marquardt, encontra-se na computação de α_k por meio da aproximação da matriz Hessiana, como é mostrado na próxima equação [70].

$$E''(\bar{w}_k)\bar{p}_k \approx \frac{E'(\bar{w}_k) + \sigma_k \bar{p}_k - E'(\bar{w}_k)}{\sigma_k} + \gamma_k \bar{p}_k \quad (3.3)$$

onde σ_k e γ_k são as constantes que regulam a aproximação da segunda derivada. O parâmetro σ_k deve ter um valor próximo de zero, uma vez que o limite da aproximação tende para o valor true (verdadeiro). O objectivo do parâmetro γ_k é garantir que a matriz Hessiana da função de erro global é positiva. Uma vez que os pesos da rede são actualizados, uma nova direcção conjugada é calculada para a próxima iteração. O procedimento é baseado na combinação da direcção de descida mais íngreme real com a direcção da pesquisa anterior, como é mostrado na equação abaixo:

$$\bar{p}_{k+1} = -E'(\bar{w}_{k+1}) + \beta_k \bar{p}_k, \beta_k = -E'(\bar{w}_1) \quad (3.4)$$

O algoritmo tenta atingir o mínimo de funções quadráticas. Porque a função de erro E é não quadrática, o algoritmo não tem necessariamente de convergir após um determinado número de iterações. Assim, durante o processo de aprendizagem, o algoritmo é reiniciado e inicia com uma direcção de busca próxima da descida mais íngreme real [70].

3.1.6 Função de erro cross-entropy para duas classes

Considerando uma RNA com uma única camada de saída y em que o valor de y representa a probabilidade posterior $P(C_1|x)$ para a classe C_1 , a probabilidade à posteriori da classe C_2 é dada por $P(C_2|x) = 1 - y$. Isto pode ser alcançado se se considerar um esquema de codificação de destino para o qual $t = 1$ se o vector de entrada pertence à classe C_1 e $t = 0$, se ele pertencer à classe C_2 . Pode-se combinar isto numa única expressão, onde a probabilidade de observar um valor-alvo é:

$$p(t | x) = y^t (1 - y)^{(1-t)} \quad (3.5)$$

Esta função é o caso particular da distribuição binomial chamada de distribuição de Bernoulli. Com esta interpretação de activações das unidades de saída, a probabilidade de observar o treino do conjunto de dados, assumindo que os pontos dos dados são extraídos de forma independente desta distribuição, é dada por:

$$\prod_n (y^n)^{t^n} (1 - y^n)^{1-t^n} \quad (3.6)$$

Como de costume, é mais conveniente minimizar o logaritmo negativo da probabilidade. Isso leva à função de erro de entropia cruzada sob a forma,

$$E = - \sum_n \{t^n \ln(y^n) + (1 - t^n) \ln(1 - y^n)\} \quad (3.7)$$

De seguida são descritas propriedades elementares da função de erro. Diferenciando a função de erro em relação a y^n obtém-se:

$$\frac{\partial E}{\partial y^n} = \frac{y^n - t^n}{y^n(1 - y^n)} \quad (3.8)$$

O mínimo absoluto da função de erro ocorre quando

$$y^n = t^n \text{ para todo o } n \quad (3.9)$$

Considerando uma rede neuronal com uma única camada de saída $y = g(a)$ em que o valor de saída é interpretado como uma probabilidade, normalmente utiliza-se a função logística como função de activação,

$$g(a) = \frac{1}{1 + e^{-a}} \quad (3.10)$$

Que tem a propriedade,

$$g'(a) = g(a) \cdot (1 - g(a)). \quad (3.11)$$

Combinando as duas equações anteriores pode-se chegar à conclusão que a derivada do erro em relação a , tem a formula simples de,

$$\delta^n = \frac{\partial E}{\partial a^n} = y^n - t^n \quad (3.12)$$

δ é a "quantidade de erro" que é retropropagado através da rede, a fim de calcular as derivadas da função de erro no que diz respeito aos pesos da rede. Note-se que tem a mesma forma obtida para a soma dos quadrados da função de erro e unidades de saída lineares. Vê-se que existe um emparelhamento natural da função de erro de saída com a função de activação da unidade que dá origem a esta forma simples das derivadas [71].

A utilização da função de activação logística também leva a simplificações ao avaliar a matriz Hessiana (matriz de segundas derivadas da função de erro). O valor mínimo da função de erro de entropia cruzada é dado por:

$$E_{min} = - \sum_n \{t^n \ln(t^n) + (1 - t^n) \ln(1 - t^n)\} \quad (3.13)$$

Para o esquema de codificação $1 - c$ a função anterior desaparece. No entanto, também é correcto usar a função de erro 3.7 quando t_n é uma variável contínua no intervalo $[0,1]$ representando assim a probabilidade do vector de entrada x_n pertencer à classe C_1 . Neste caso, o valor mínimo do erro não precisa desaparecer, e por isso é conveniente subtrair esse valor à função de erro original de forma a obter o erro de modificação dado por,

$$E_{min} = - \sum_n \left\{ t^n \ln \left(\frac{y^n}{t^n} \right) + (1 - t^n) \ln \left(\frac{1 - y^n}{1 - t^n} \right) \right\} \quad (3.14)$$

Uma vez que a equação 3.14 é independente da saída da RNA isso não afecta a localização do mínimo e assim não tem nenhum efeito sobre a formação da rede. O erro modificado tem sempre o seu valor mínimo em 0, independentemente do conjunto de treino específico.

3.1.7 Weight Decay

Um processo que permite evitar o “overfitting” (sobre ajustamento ao conjunto de treino por excesso de treino perdendo a capacidade de generalização) é o de utilizar a técnica de Weight Decay (WD), uma técnica que procura evitar que os pesos se afastem significativamente de zero. Além dos pesos serem actualizados do modo usual, são também diminuídos de uma fracção λ do seu valor anterior. Assim, enquanto o processo de treino altera os pesos de modo a diminuir o erro, o WD procura que eles se mantenham próximo de zero, evitando o “overfitting”. Em termos de parâmetros efectivos esta técnica impede que os pesos participem completamente no processo de modelização. A técnica de WD favorece fortemente a ocorrência de muitos pesos de pequena magnitude. O parâmetro WD pode ser definido como:

$$\Delta \omega_{ij}(t + 1) = n \delta_i o_j - \lambda \omega_{ij}(t) \quad (3.15)$$

3.1.8 Registo cronológico do desenvolvimento das redes neuronais

Década de 40, a estrutura do neurónio artificial de McCulloch e Pitts, uma proposta para a descrição matemática de cálculos lógicos com base nas estruturas neuronais biológicas desencadeou todo o processo de investigação nesta área [24]. Em 1949, Donald Hebb propôs um método de proporcionar capacidade de aprendizagem às redes neuronais artificiais. Em 1956, na primeira Conferência Internacional de Inteligência Artificial, foi apresentado um modelo de rede neuronal artificial pelo pesquisador Nathaniel Rochester. Década de 60, o Perceptron de Rosenblatt, em que um modelo neuronal não linear e adaptativo permitiu expressar os processamentos representados no modelo de McCulloch e Pitts e incorporar um algoritmo bastante eficaz [24]. Em 1969, Marvin Minsky e Seymour Papert publicaram o “Perceptrons”, onde expõem as limitações do modelo de Rosenblatt, provando que tais redes não são

capazes de resolver uma ampla classe de problemas devido às restrições de representação. O impacto desta publicação foi devastador, praticamente desaparecendo o interesse em RNAs nos anos seguintes. Década de 80, em 1982, John Hopfield apresentou um sistema neuronal artificial capaz de armazenar nas interligações entre os neurónios informações complexas, como imagens por exemplo. Parte das dúvidas até então existentes em relação ao processo dinâmico executado por certas RNAs são dissipadas. Em 1986, Rumelhart, Hinton, Williams e McClelland desenvolveram e publicaram o algoritmo de “back-propagation” que permite treinar redes neuronais em multicamadas, que permitiram resolver os problemas não lineares apresentados por, Marvin Minsky e Seymour Papert. Década de 90, as RNAs tiveram uma explosão exponencial de aplicações e desenvolvimento de modelos. A partir daí, consolidam-se as RNAs como parte integrante do estudo da Inteligência Artificial propriamente dita.

3.1.9 Aplicações das Redes Neurais Artificiais

São inúmeras as aplicações de redes neuronais artificiais nos mais diversos campos, seria morosa a enumeração de todos pelo que serão enunciados apenas alguns: análise de séries temporais, análise e processamento de sinais, reconhecimento de texto, imagem e voz, reconhecimento de padrões, controlo de processos, classificação de dados, análise de mercados financeiros, segurança, pesquisa de mercado, transportes, banca e avaliação de crédito, identificação de fraude de cartão de crédito, aplicações climáticas, indústria automóvel, indústria de projecção petrolífera e gás, indústria aeroespacial, robótica: controlo da trajectória e sistemas de visão electrónica, telecomunicações, economia, gestão e seguros, medicina e biomédica. Alguns exemplos concretos associados à engenharia biomédica: análise de células cancerígenas, desenho de próteses, optimização do tempo de transplante, análise de E.E.G. e E.C.G, redução de despesas hospitalares, apoio à decisão em neurocirurgia através da Internet (o usuário fornece os dados do paciente com traumatismo crânio-encefálico e a rede neuronal responde ao prognóstico e auxílio sobre decisão de se fazer ou não a cirurgia), diagnóstico de enfarte agudo do miocárdio, análise de batimentos cardíacos, diagnóstico médico da Tuberculose, detecção de eventos epilépticos, análise de variáveis fisiológicas e clínicas de pacientes em estado crítico, análise de variáveis obtidas pelos exames ecocardiográficos doppler de pacientes com insuficiência cardíaca.

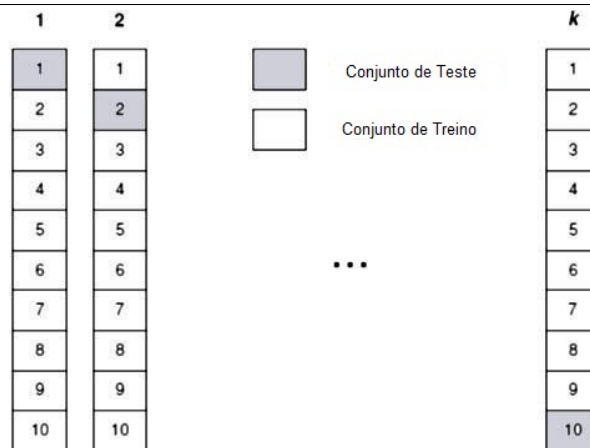
3.1.9.1 Condições para a aplicação da RNA

A RNA pode ser aplicada quando existe um conjunto de dados exemplo do problema que serão transformados em vectores de entrada/saída. Estes dados devem ser representativos do problema. O seu número deve ser grande para treinar a rede (pelo menos várias vezes o número de pesos da rede) [69].

3.1.10 Validação cruzada deixando um de fora (leave - one - out - cross - validation)

Neste tipo de validação o conjunto de teste tem apenas um exemplo e o conjunto de treino é composto por $k - 1$ exemplos. A principal vantagem deste método é a de utilizar o maior número de instâncias possíveis para treinar o classificador, o que pode aumentar as probabilidades do classificador induzido se tornar mais preciso (Confrontar Figura 3.6). Mais ainda, os conjuntos de teste definidos nas diferentes instâncias são mutuamente exclusivos e cobrem completamente o conjunto de dados. A desvantagem mais evidente está ligada ao esforço computacional que é necessário para repetir o procedimento k vezes. Pode-se ainda apontar como desvantagem o facto de o conjunto de teste ser constituído por apenas um exemplo, originando assim uma variância elevada da estimativa do desempenho medido [72].

Figura 3.6 Processo de validação cruzada deixando um de fora.



Cada um dos quadrados representa um elemento.

3.2 Regressão Logística

A regressão logística tem vindo a ser utilizada nas mais diversas áreas da ciência. Este método, assim como as regressões lineares e múltiplas, estuda a relação entre uma variável resposta e uma ou mais variáveis independentes. A diferença entre estas técnicas de regressão deve-se ao facto de que na regressão logística as variáveis dependentes estão dispostas em categorias, enquanto na regressão linear estas variáveis são dados contínuos ou discretos. Outra diferença é que na regressão logística a resposta é expressa por meio de uma probabilidade de ocorrência, enquanto que na regressão simples obtém-se um valor numérico [73].

3.2.1 Regressão Logística Binária

Uma variável binária é aquela que aceita apenas dois níveis de resposta, como sim ou não. Existem vários tipos de estudos que se podem analisar com modelos logísticos. Estes incluem epidemiologia,

experiências clínicas, pesquisa de mercado, entre outras. O modelo logístico é projectado para ajustar dados binários, quando é assumido que y representa uma distribuição de probabilidades, ou quando é simplesmente expresso como uma medida binária que se está a tentar prever. Na Regressão Logística Binária temos que a variável resposta possui apenas duas categorias, ou seja, natureza binária ou dicotómica, e apenas uma variável independente envolvida. Para analisar $P(x)$, que é a probabilidade de um acontecimento ocorrer, utilizam-se as observações independentes x_1, x_2, \dots, x_n . Neste contexto, é razoável assumir, como suposição inicial, que $P(x)$ é uma função monótona com valores entre zero e um, quando x varia na recta real, ou seja, $P(x)$ é uma função de distribuição de probabilidade. Como $P(x)$ varia entre zero e um, uma representação linear simples para $P(x)$ sobre todos os valores de x não é adequada, então considera-se a transformação logística de $P(x)$ sob a forma linear [74].

$$\ln(P(x)/(1 - P(x))) = g(x)$$

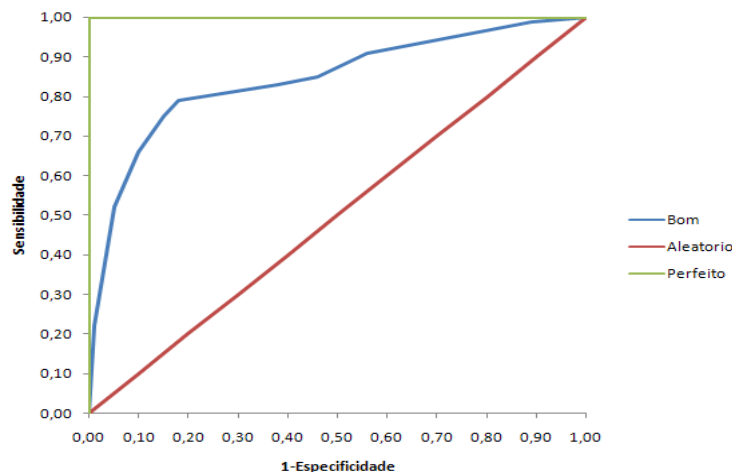
$$\text{onde, } g(x) = \beta_0 + \beta_1 x \quad (x) = (\exp(\beta_0 + \beta_1 x))/(1 + \exp(\beta_0 + \beta_1 x)) \quad (3.16)$$

onde, é necessário que $\beta_1 < 0$ para que P seja crescente e que $\beta_1 > 0$ para que P seja decrescente. Quando x tende para infinito, $P(x)$ tende para zero quando $\beta_1 < 0$ e tende para um quando $\beta_1 > 0$. Assim, dessa forma, define-se a função de ligação necessária ao modelo. Caso $\beta_1 = 0$, a variável de resposta y é independente da variável. Na equação tem-se $\beta_0 = 0$ e $\beta_1 = -1$ então $P(x)$ é chamada de função de distribuição logística. E conseqüentemente, sua função de densidade de probabilidade é obtida com a derivação de esta função [74].

3.3 Métodos de apoio à decisão

3.3.1 Curvas ROC

A curva ROC foi desenvolvida por engenheiros eléctricos e engenheiros de sistemas de radar durante a Segunda Guerra Mundial para detectar objectos inimigos em campos de batalha, também conhecida como teoria de detecção de sinais. A análise ROC tem sido utilizada ao longo dos tempos em medicina, radiologia e psicologia e mais recentemente na aprendizagem de máquina. Como o resultado de sistemas de classificação em classes geralmente são contínuos, ou seja, produzem um valor situado dentro de um determinado intervalo contínuo, como $[0;1]$, é necessário definir um ponto de corte, ou um limiar de decisão, para se classificar e contabilizar o número de predições positivas e negativas (como nos diagnósticos verdadeiros e falsos no caso de ocorrência de uma patologia). Como este limiar pode ser seleccionado arbitrariamente, a melhor prática para se comparar o desempenho de diversos sistemas é estudar o efeito

Figura 3.7 Curva Roc.

de selecção de diversos limiares sobre a saída dos dados. Para cada ponto de corte são calculados valores de sensibilidade e especificidade, que podem ser dispostos num gráfico denominado curva ROC, que apresenta no eixo das ordenadas os valores de sensibilidade e nas abcissas o complemento da especificidade, ou seja, o valor (1-especificidade). Um classificador perfeito corresponderia a uma linha horizontal no topo do gráfico (Figura 3.7), porém esta dificilmente será alcançada. Na prática, curvas consideradas boas estarão entre a linha diagonal e a linha perfeita, onde quanto maior a distância em relação à linha diagonal, melhor será o sistema. A linha diagonal indica uma classificação aleatória, ou seja, um sistema que aleatoriamente selecciona saídas como positivas ou negativas. Definitivamente, não é o tipo de sistema mais confiável possível. No entanto, um sistema cuja curva ROC esteja localizada abaixo da diagonal ainda pode ser convertido num bom sistema, para isso, basta inverter as suas saídas e então a sua curva também será invertida [75].

Uma medida padrão para a comparação de sistemas é a área sob a curva (AROC), que pode ser obtida por métodos de integração numérica, como por exemplo, o método dos trapézios. Teoricamente, quanto maior a AROC, melhor o sistema. Outras medidas como a sensibilidade, especificidade e precisão podem ser obtidas pela chamada tabela de contingência (ou matriz de confusão). O seu funcionamento é simples: consideramos valores positivos, os que o sistema julgou positivos como verdadeiros positivos (acerto), valores positivos os que o sistema julgou negativos como falsos negativos (erro), valores negativos os que o sistema julgou negativos como verdadeiros negativos (acerto), e valores negativos os que o sistema julgou positivos como falsos positivos (erro). As medidas de Precisão, Sensibilidade e Especificidade obtidas através da Tabela 3.2 são descritas de seguida [75].

Precisão é a proporção de predições correctas, sem levar em consideração o que é positivo e o que é negativo. Esta medida é altamente susceptível a desbalanceamentos do conjunto de dados e pode facilmente induzir a uma conclusão errada sobre o desempenho do sistema.

$$Precisão = \frac{Total\ de\ acertos}{Total\ de\ dados\ no\ conjunto} = \frac{VP + VN}{P + N} \quad (3.17)$$

Sensibilidade é a proporção de verdadeiros positivos, isto é, traduz a capacidade do sistema em prever correctamente a condição para os casos que realmente são DA.

$$Sensibilidade = \frac{Acertos\ Positivos}{Total\ de\ Positivos} = \frac{VP}{VP + FN} \quad (3.18)$$

Especificidade é a proporção de verdadeiros negativos, demonstrando assim a capacidade do sistema em prever correctamente a ausência da condição para os casos que realmente não são DA [75].

$$Especificidade = \frac{Acertos\ Negativos}{Total\ de\ Negativos} = \frac{VN}{VN + FP} \quad (3.19)$$

Tabela 3.2 Matriz de Confusão / tabela de contingência.

		Valor Verdadeiro (Confirmado por análise)	
		Positivo	Negativo
Valor Previsto (predito pelo teste)	Positivos	VP Verdadeiro Positivo	FP Falso Positivo
	Negativos	FN Falso Negativo	VN Verdadeiro Negativo

3.3.2 P-Valores

Quando se colhe uma amostra da população, não se sabe com certeza se alguma estatística (como a média) dessa amostra corresponde realmente à população. Considerando que a amostra cobre somente uma fracção de um todo, não se espera, em geral, que a estatística da amostra corresponda exactamente à da população. Os testes de hipóteses são usados para verificar se as diferenças entre os valores da amostra e os valores da população são devidos a um mero acaso. Primeiro elabora-se uma hipótese sobre a população da qual a amostra foi retirada. Esta é a chamada hipótese nula, H_0 , porque propõe que não exista diferença entre a amostra e a população, no aspecto que está a ser considerado. Em contrapartida, a hipótese alternativa, H_1 é formulada para testar a hipótese contrária à hipótese nula.

$$H_0 : \theta = \theta_0 \text{ (hipótese nula);} \quad (3.20)$$

$$H_1 : \theta \neq \theta_0 \text{ (hipótese alternativa);} \quad (3.21)$$

Os testes podem rejeitar ou aceitar a hipótese nula. Há dois possíveis tipos de erros quando realizamos um teste estatístico para aceitar ou rejeitar H_0 . Erro do tipo I: é o erro ao rejeitar H_0 quando, na realidade, H_0 é verdadeira. A probabilidade de cometer este erro do tipo I é designada por α (nível de significância). O erro do tipo I equivale a concluir que o tratamento é eficaz quando na verdade ele não é. Erro do tipo II : é o erro ao aceitar H_0 quando, na realidade, H_0 é falsa. A probabilidade de cometer este erro do tipo II é designada por β .

Tabela 3.3 Tomada de decisão de um teste de Hipóteses.

Decisão	H_0 VERDADEIRA	H_0 FALSA
Aceitar H_0	Decisão correta Probabilidade = $1 - \alpha$	Erro do tipo II Probabilidade = β
Rejeitar H_0	Erro do tipo I Probabilidade = α	Decisão correta Probabilidade = $1 - \beta$

Num teste de hipóteses é obviamente desejável que se reduza ao mínimo as probabilidades α e β dos dois tipos de erros. Porém, a diminuição de se ter um erro implica o aumento probabilidade de se ter um outro erro. Em geral, escolhe-se pela diminuição do erro tipo I. A redução simultânea dos erros poderá ser alcançada pelo aumento do tamanho da amostra. Para um teste de hipóteses, determina-se a probabilidade máxima de aceitar o erro tipo I. Essa probabilidade máxima é chamada de nível de significância, e pode ser estipulada de acordo com o pesquisador. Em geral, estipula-se um nível de 5%. O valor da probabilidade de se obter o efeito observado, dado que a hipótese nula é verdadeira, é chamado de p -valor. Se o valor do p -valor for menor que o nível de significância estipulado, assume-se o erro tipo I e rejeita-se a hipótese nula. Ao contrário, se o p -valor for maior, não é assumido o erro tipo I e aceita-se a hipótese nula. O p -valor não é a probabilidade de que o efeito observado tenha ocorrido por acaso, mas sim a probabilidade de obter o efeito observado (ou um valor menos provável) quando a hipótese é verdadeira. Outra interpretação falsa é a de que um p -valor de 0.005 significa que tem um efeito maior do que um p -valor de, digamos 0.05, por exemplo. Embora isto possa ser verdade, os p -valores não demonstram isto. Vale a pena ressaltar que um teste de hipóteses não prova nem desaprova hipóteses. Na verdade, um teste de hipóteses é uma ferramenta para tomar decisões, mas está sujeita a erros. Em resumo, um teste de hipóteses é feito inicialmente pela Formulação de H_0 e de H_1 ; de seguida ocorre a fixação de α e consequente determinação dos valores da variável X que levam à aceitação de H_0 ; tais valores de X definem a chamada região de aceitação de H_0 e permitem criar a região de rejeição;

No final é apresentada a amostra X de tamanho n ao teste e ocorre a observação de X e a comparação dos resultados com a região de aceitação. Se o X observado corresponder a um dos valores de X da região de aceitação, aceita-se H_0 ; caso contrário, rejeita-se H_0 [76].

3.3.2.1 Teste de Kolmogorov-Smirnov

Analisa se os dados da amostra foram extraídos de uma população com uma distribuição peculiar de frequências, como a distribuição Normal. Se aceitar a hipótese nula H_0 a distribuição de parâmetros é normal, se rejeitar H_0 não segue uma distribuição Normal.

3.3.2.2 Teste de Levene

Este teste é usado quando há várias amostras, e verifica se é plausível que as amostras tenham, ou não, todas a mesma variância. Isto é, testa a homoscedasticidade das amostras. Se aceitar a hipótese nula H_0 as distribuições são homogêneas, se rejeitar H_0 as distribuições não são homogêneas.

3.3.2.3 Teste de Kruskal-Wallis

O teste de Kruskal-Wallis realiza uma análise unidireccional da variância por postos (em homenagem a William Kruskal e Wallis W. Allen) é um método não-paramétrico para testar a igualdade das medianas da população entre os grupos. Ele é idêntico a uma análise unidireccional de variância. Como tal, é um equivalente não paramétrico de um factor inteiramente projectado da análise de variância. Supõe-se que diversas populações têm a mesma forma e dispersão para que hipótese acima seja aplicável, porque as diferenças na forma ou na dispersão levariam à rejeição da hipótese nula. Os valores das várias amostras independentes aleatórias são obrigados a estar na escala ordinal. As várias amostras são primeiramente vistas como uma matriz de valores, e cada valor no grupo combinado é distribuído do menor para o maior. Se a hipótese nula for verdadeira, a média das classificações para cada grupo de amostra deve ser aproximadamente igual. A estatística de teste calculada é designada H e é baseada na soma de cada um dos postos em cada uma das várias amostras aleatórias, como se segue:

$$H = \left(\frac{12}{N(N+1)} \right) \left[\sum \frac{(R_j^2)}{n_j} \right]; \quad (3.22)$$

em que N é o tamanho das amostras combinadas; R_j é a soma dos postos de cada uma das amostras; e n_j é o número de casos por cada grupo. Como é um método não-paramétrico, o teste de Kruskal-Wallis não assume uma população normal, ao contrário das análises unidireccionais de variâncias análogas. No

entanto, o teste não assume uma distribuição idêntica em forma e escala para cada grupo, excepto no caso das diferenças de medianas [77].

3.3.3 Análise de componentes principais (PCA)

Introduzida por Karl Pearson em 1901, a técnica denominada análise de componentes principais, popularmente chamada PCA (*principal components analysis*), está fundamentada no artigo [78]. A PCA é um método estatístico multivariado, não-paramétrico, não-supervisionado e linear, baseado na transformação Karhunen-Loève, usado para modelos de classificação no reconhecimento de padrões [79]. A PCA tem como objectivo principal explicar a estrutura de variância e co-variância de um vector aleatório, composto de p -variáveis aleatórias, através da construção de combinações lineares das variáveis originais. As componentes principais são as combinações lineares e não são correlacionadas entre si. Se tivermos p -variáveis originais que representam um dado então será possível obter p componentes principais. No entanto, o objectivo é obter a redução do número de variáveis a serem avaliadas e a interpretação das combinações lineares construídas, ou seja, a informação contida nas p -variáveis originais é substituída pela informação contida em k componentes principais não correlacionadas, sendo $k < p$. Sendo assim, a variabilidade do vector aleatório composto das p -variáveis original é aproximado pela variabilidade do vector aleatório que deverá conter as k componentes principais. Desta forma, a qualidade da aproximação dependerá do número de componentes mantidas no sistema e pode ser medida através da avaliação da proporção de variância total explicada por essas componentes escolhidas. A PCA é uma forma de identificar padrões em dados, e de expressá-los de tal forma a destacar as suas similaridades e diferenças. Devido à alta dimensionalidade dos dados, a sua representação gráfica nem sempre é possível, daí PCA torna-se uma poderosa ferramenta para analisar tais dados. Uma outra vantagem é que uma vez encontrados padrões nos dados, comprime-se os dados reduzindo o número de dimensões sem muita perda de informação. Uma vez determinadas as componentes principais, os seus valores numéricos, podem ser calculados para cada elemento da amostra. Deste modo, os valores de cada componente podem ser analisados, fazendo uso de técnicas estatísticas como análise de variância e análise de regressão, entre outras, de forma a permitir eliminar as componentes similares que podem levar a erros durante a classificação [79].

Descreve-se seguidamente o algoritmo da Análise de Componentes Principais:

1. Transformação da matriz dos dados originais. Os dados iniciais são transformados numa matriz x através da operação de redução de variáveis iniciais:

$$x_{ij} = \frac{1}{\sqrt{n}} \cdot \frac{q_{ij} - \bar{q}_j}{s_j} \quad (3.23)$$

onde q_{ij} representa o valor da j -ésima variável do conjunto de dados de i . \bar{q}_j é a média aritmética dos valores da variável j , s_j é o desvio padrão da variável j . O coeficiente $\frac{1}{\sqrt{n}}$ tem por objectivo fazer coincidir a matriz $X^T X$ com a matriz covariância.

2. Cálculo da matriz de correlações r , cujo elemento genérico é dado por:

$$r_{jj'} = \frac{\sum_{i=1}^n x_{ij}x_{i'j'}}{n \cdot s_j \cdot s_{j'}} = \frac{1}{n} \sum_{i=1}^n \frac{(q_{ij} - \bar{q}_j)(q_{i'j'} - \bar{q}_{j'})}{s_j \cdot s_{j'}} \quad (3.24)$$

3. Diagonalização da matriz de correlações de que resultam p valores próprios (λ_α) e p vectores próprios $u_{\alpha*}$
4. Cálculo das coordenadas das variáveis nos eixos factoriais, dadas por:

$$f_{j\alpha} = \sum_{i=1}^n x_{ij}v_{i\alpha} \quad (3.25)$$

5. Cálculo das projecções das variáveis nos eixos factoriais, dadas por:

$$w_{i\alpha} = \sum_{j=1}^p x_{ij}u_{j\alpha} \quad (3.26)$$

6. Selecção da dimensão do sub-espço, cuja inércia acumulada explique uma percentagem suficiente da inércia total.
7. Interpretação dos resultados.

Classificação dos Sinais EEG

Este capítulo inicia-se com uma descrição da base de dados utilizada neste estudo. Seguidamente, explica-se detalhadamente o pré-processamento efectuado aos sinais de EEG. E por fim, faz-se a descrição e a explicação de cada uma das metodologias aplicadas (WT, STFT e DWT), com vista à discriminação dos sinais de EEG entre pacientes com DA e sujeitos de controlo. Os resultados da discriminação, são apresentados e avaliados no final de cada uma das metodologias.

4.1 Recolha e Pré-Processamento do Sinal de EEG

4.1.1 Selecção dos pacientes e dos sujeitos de controlo

Trinta e quatro indivíduos participaram no estudo, 14 sujeitos de controlo e 20 pacientes com probabilidade de terem DA. Todas as gravações de EEG foram realizadas no Hospital Universitário Pio de Ró Hortega (Valladolid, Espanha). Catorze sujeitos (sendo 8 homens e 6 mulheres, com idade = $75,5 \pm 1,34$ (61-87) anos), participaram no estudo como grupo de controlo, encontravam-se cognitivamente normais, sem histórico de distúrbios neurológicos ou psiquiátricos. Vinte pacientes com probabilidade de DA (sendo 5 homens e 15 mulheres com a idade = $78,20 \pm 7,49$ (58-91) anos), foram incluídos no estudo, todos eles foram submetidos a uma avaliação clínica cuidadosa, que inclui histórico clínico, exame físico, neurológico e cerebral.

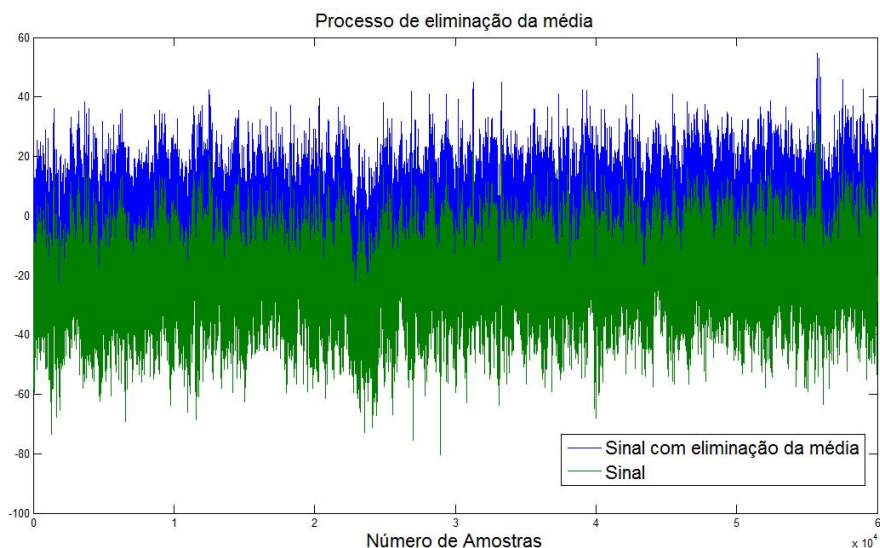
4.1.2 Gravação dos Sinais de EEG

Os EEGs foram realizados a partir do couro cabeludo, através de um sistema internacional 10-20 de 19 localizações (eléctrodos Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz e Pz) usando um electroencefalógrafo digital Xltek de Natus Médico Inc. com uma frequência de amostragem de 200 Hz (ver Figura 1.1). As gravações foram feitas com os indivíduos num estado de relaxamento e sob a condição de olhos fechados, a fim de se obter o maior número de dados de EEG, livres de artefactos.

4.1.3 Pré-Processamento dos sinais de EEG

Os sinais EEG foram carregados no Software Matlab sob a forma de matrizes de 60.000x19. Por eléctrodo, procedeu-se à eliminação da média, isto é, os sinais dos eléctrodos sofreram uma deslocação no eixo das ordenadas pela subtracção de cada um dos seus valores, pelo valor médio do sinal, conforme mostra a Figura 4.1.

Figura 4.1 Processo de eliminação da média por eléctrodo.



Todas as gravações foram digitalmente filtradas através um filtro FIR passa banda de janela Hamming de ordem 300 com uma frequência de passagem de 1 Hz a 40 Hz. A Figura 4.2 ilustra 5 segundos do sinal de EEG antes de ser filtrado e a Figura 4.3 ilustra os mesmos 5 segundos do sinal, depois de ser filtrado.

Figura 4.2 5 segundos de Sinal EEG.

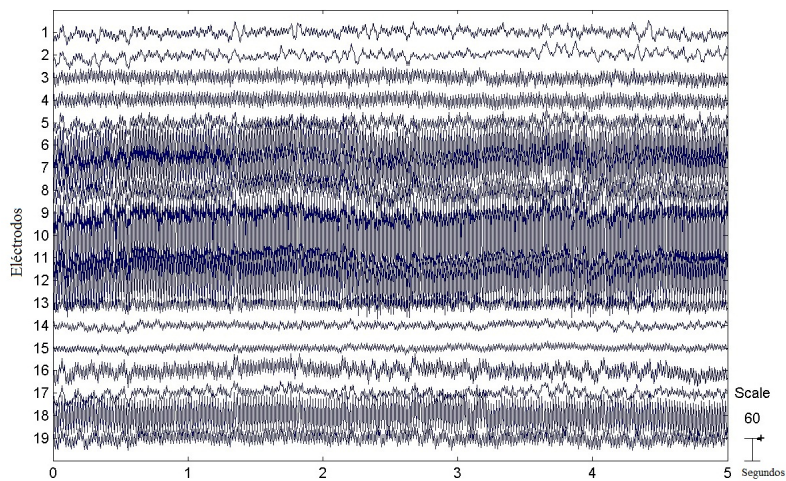
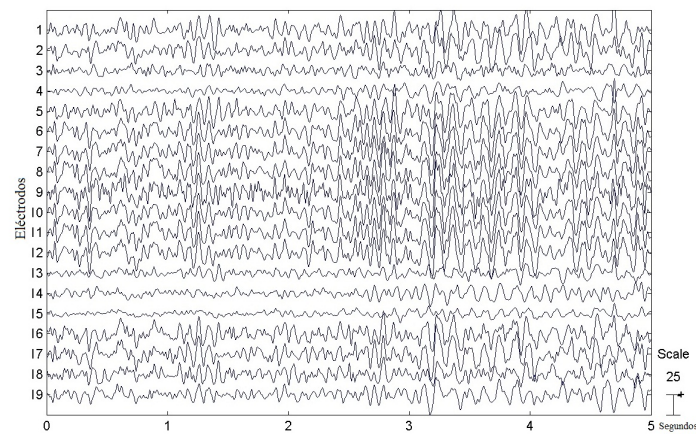


Figura 4.3 5 segundos de Sinal EEG filtrado.

Os EEGs foram inspeccionados visualmente por um médico especialista para verificar o movimento dos olhos e outros artefactos. Assim, apenas os dados de EEG livre de artefactos de movimento e com o mínimo de electromiografia (EMG) foram seleccionados para este estudo.

De seguida, os EEGs foram organizados em segmentos de cinco segundos livres de artefactos (1000 amostras). No final ficou-se com um número médio de $31,2 \pm 8,1$ (média \pm desvio padrão(SD)) segmentos livres de artefactos para cada um dos eléctrodos de cada indivíduo.

4.2 Classificação dos sinais EEG utilizando a *PSD* baseada na *WT*

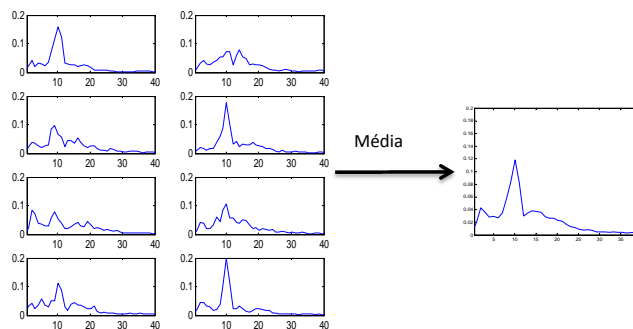
A técnica de *PSD* obtida através da *WT* foi utilizada para a extracção de características dos sinais EEG. Por cada origem diferente de características procedeu-se à classificação desses sinais. Os *p*-valores por indivíduo e por eléctrodo, os parâmetros obtidos pela curva ROC, bem como os biplots da técnica de PCA, permitiram uma visão mais clara das combinações de parâmetros que possibilitaram uma maior eficiência da RNA e da RL, em discriminar as suas entradas. Foram criados conjuntos de treino e de teste de modo a proceder à classificação dos sinais EEG através das RNAs e da RL. Avaliaram-se estas duas técnicas de classificação pelo processo de classificação por validação cruzada deixando um de fora, respectivamente, utilizando o conjunto total de casos envolvidos em cada uma das abordagens à metodologia.

4.2.1 Processamento do sinal EEG e extracção de características

A *PSD* do sinal baseada na *WT* foi calculada até ao nível de decomposição 7 da *WPT*, o que permitiu obter o espectro de *WT* do sinal desde 1Hz até 40Hz. Esse intervalo contém as frequências clássicas principais de um sinal de EEG. Foi determinada a média das várias *PSDs* ao longo dos sucessivos segmentos de 5s de duração do sinal por eléctrodo (PSD_n por eléctrodo), e de seguida a média das várias PSD_n por

eléctrodo (PSD_n por indivíduo) (ver exemplo, Figura 4.4). No cálculo das várias PSD s utilizaram-se as Wavelets Db8, Sym8, Bior3.5 e Coif3. A PSD foi normalizada no intervalo de 1 a 40Hz (equação 2.18).

Figura 4.4 Exemplo de um Processo de cálculo da média das PSD_n de 5s de duração (PSD_n por eléctrodo).



4.2.2 Utilização dos parâmetros RP e rácios espectrais

Foram calculadas 10 características (RP em delta, teta, alfa, beta1, beta2, gama, bem como os rácios espectrais associados r_1, r_2, r_3, r_4), de acordo com as equações 2.26 à 2.35, de forma a poder resumir a dinâmica do sinal EEG a partir da PSD_n .

4.2.2.1 Análise estatística / selecção de características

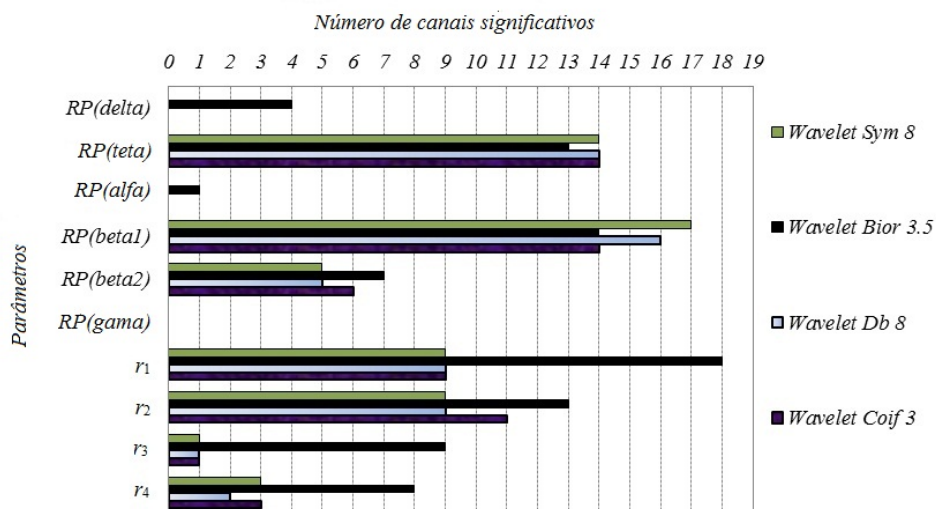
Foi realizada a análise descritiva das distribuições de parâmetros, extraídos da PSD_n por indivíduo, dos EEGs de pacientes com DA e sujeitos de controlo. Analisou-se a normalidade das distribuições com o teste de KS e a homocedasticidade pelo teste de Levene (Tabela 4.1). Utilizou-se aleatoriamente a WT Bior 3.5 para esse fim.

Tabela 4.1 Teste de KS e teste de Levene das distribuições das RP s e dos rácios espectrais - por indivíduo.

Parâmetros	Teste KS		Teste Levene	
	Teste da Hipótese nula	p -valor	Teste da Hipótese nula	p -valor
$RP(delta)$	1	0,008	1	0
$RP(teta)$	1	0	0	0,317
$RP(alfa)$	1	0,049	0	0,559
$RP(beta1)$	1	0,004	0	0,864
$RP(beta2)$	1	0,001	0	0,579
$RP(gama)$	0	0,932	0	0,662
r_1	1	0,001	0	0,624
r_2	1	0,001	0	0,518
r_3	1	0,001	1	0,029
r_4	1	0,004	0	0,079

No caso do teste de KS, as distribuições de parâmetros em geral rejeitaram a hipótese nula (normalidade da distribuição dos parâmetros), isto é, os parâmetros não apresentaram uma distribuição normal, porque no cômputo geral, a resposta à hipótese nula para a normalidade foi negativa (1). Por outro lado, no teste de Levene, as distribuições de parâmetros em geral aceitaram a hipótese nula (homocedasticidade da distribuição dos parâmetros) dado que a resposta da grande maioria das distribuições de parâmetros à hipótese nula foi positiva (0) o que leva a dizer que as distribuições são homogêneas. Assim, concluiu-se que as distribuições não atenderam à hipótese dos testes paramétricos. Por conseguinte, estudaram-se as diferenças entre os grupos por meio do teste de Kruskal-Wallis que é um teste não-paramétrico. De forma a seleccionar a Wavelet que se ajustou melhor à dinâmica do sinal de EEG nesta abordagem, procedeu-se à comparação entre os valores dos parâmetros dos pacientes com DA e dos sujeitos de controlo extraídos da PSD_n por eléctrodo, pelo teste de Kruskal-Wallis com a correcção de Bonferroni para um $p < 0.05$, o que equivale a dizer que se verificaram as diferenças com um $p < 0.05/19$ eléctrodos = $p < 0.0026$, por eléctrodo. Apresenta-se na Figura 4.5 o gráfico que permitiu realizar a selecção da Wavelet.

Figura 4.5 Número de canais significativos para os processamentos envolvidos neste estudo com p -valores < 0.05 obtidos através do teste de Kruskal-Wallis com correcção de Bonferroni.



Pode ser observado na Figura 4.5 que a PSD de WT Bior 3.5 forneceu mais canais significativos (eléctrodos onde existem diferenças significativas ($p < 0.05$ com correcção de Bonferroni) entre pacientes com DA e sujeitos de controlo) em todos os rácios espectrais (r_1 , r_2 , r_3 , r_4) do que as restantes WT s. A WT Bior 3.5 exibiu também parâmetros significativos em alguns canais, como no caso da $RP(alfa)$ e da $RP(delta)$, onde nenhum outro processamento exibiu. As restantes RP s tiveram os mesmos canais significativos, com excepção do parâmetro $RP(beta1)$ onde a WT Sym8 apresentou mais. Finalmente, nenhum processamento mostrou canais significativos em $RP(gama)$. Em suma, para a generalidade das RP s e para os rácios espectrais associados, a WT Bior 3.5 teve mais capacidade de identificar e destacar

a actividade da DA do que as restantes WT's. Seleccionou-se por isso a Wavelet Bior 3.5 para classificar os sinais EEG nesta abordagem.

Figura 4.6 Diagramas de caixa que apresentam a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais obtidos pela PSD WT Bior 3.5 .

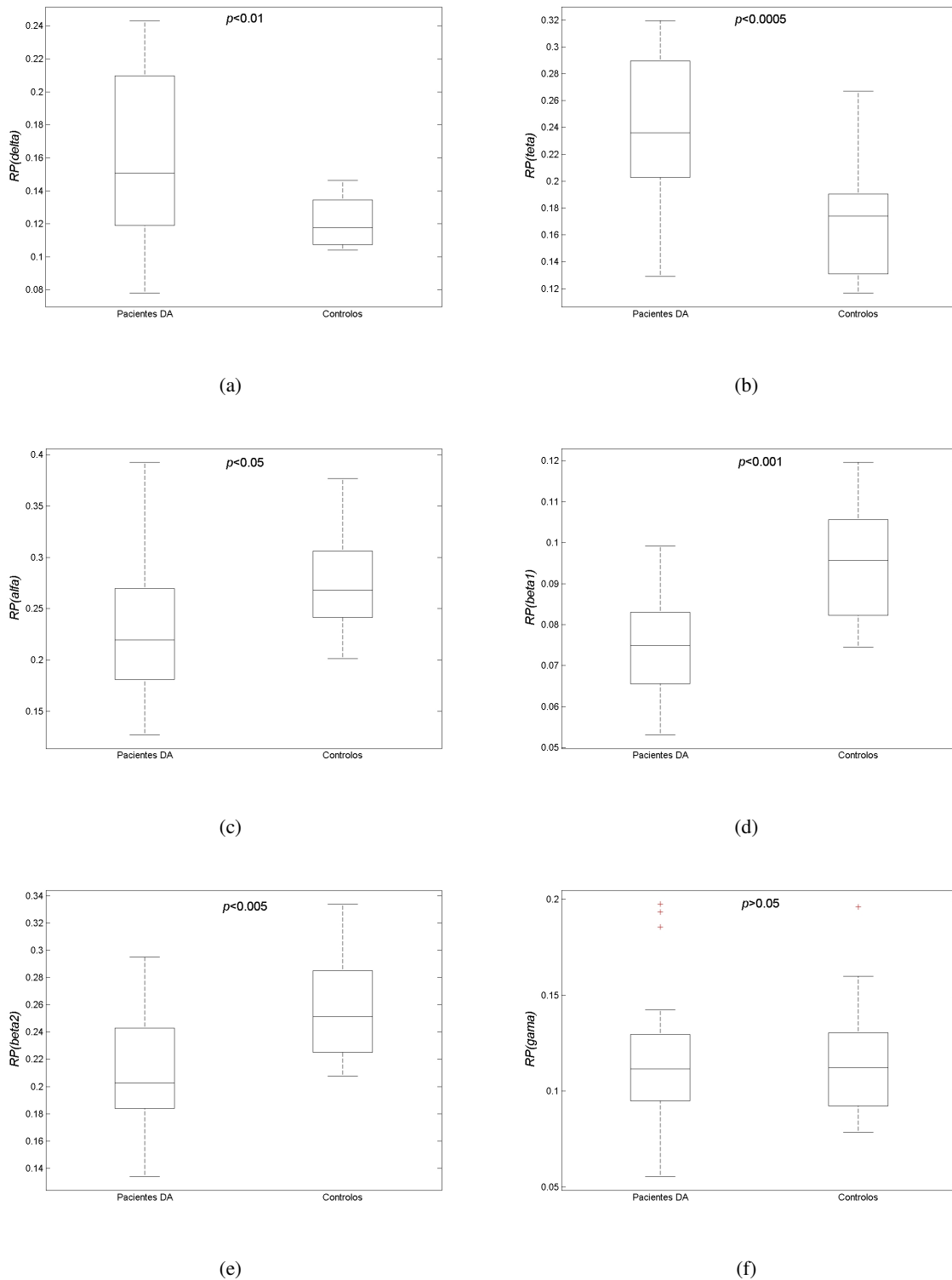
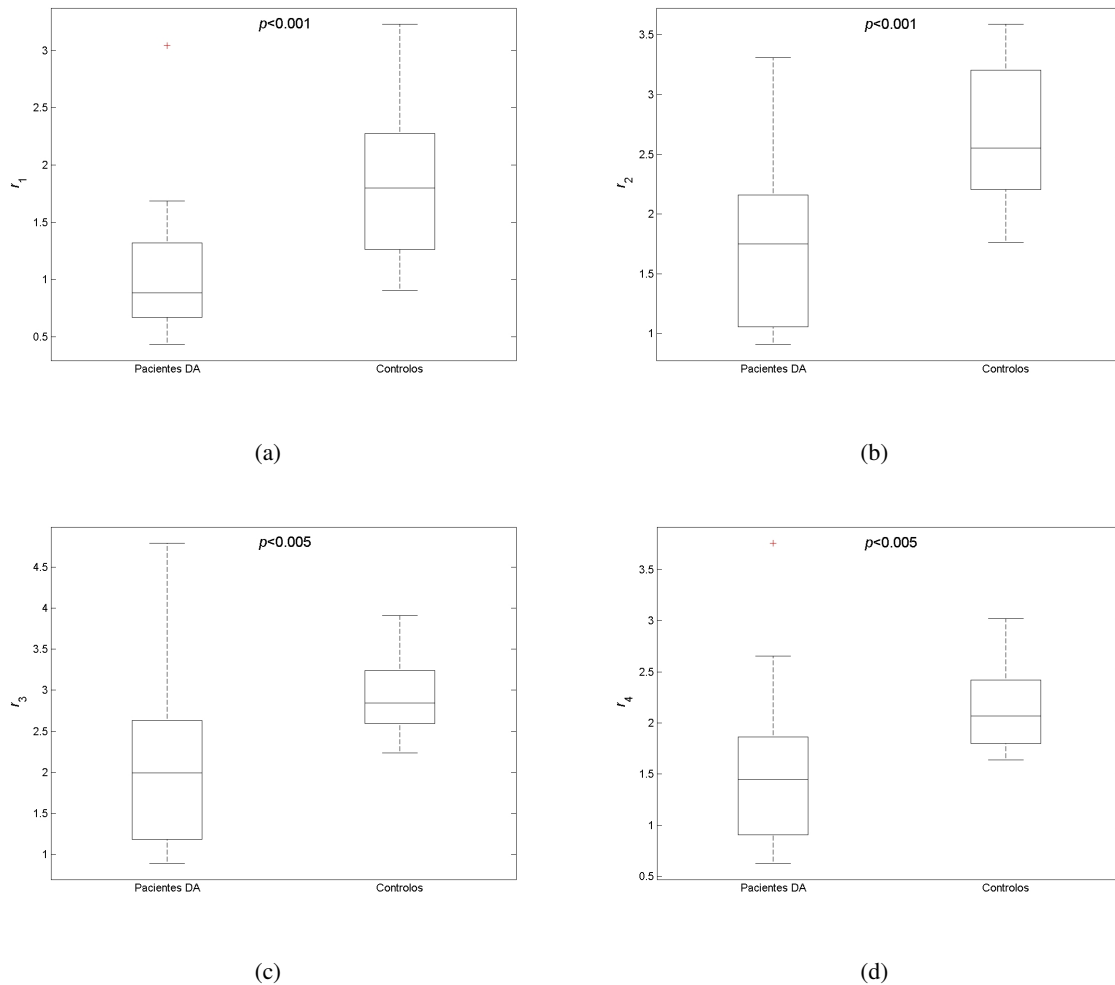


Figura 4.7 Diagramas de caixa que mostram a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rcios espectrais obtidos pela PSD_{WT} Bior 3.5 - Continuacao.



No sentido de verificar as distribuices dos parmetros, obtidos pela PSD_n por indivduo, utilizaram-se os diagramas de caixa (Figura 4.6 e Figura 4.7) para de uma forma mais resumida, visual e intuitiva se poder verificar as diferenas entre as distribuices de parmetros obtidas para os pacientes com DA e para o grupo de controlo. Os valores ordenados dos dados foram divididos em quatro partes, usando uma caixa com trs linhas horizontais. A mediana da amostra (Teste de Kruskal-Wallis) foi exibida como uma linha situada aproximadamente no meio da caixa, onde os 25 e 75 percentis da amostra se situam na parte inferior e superior da caixa, respectivamente. Finalmente, houve duas linhas de discrepncia que se desenvolveram a partir de cada extremidade da caixa at ao mximo de 1,5 vezes do intervalo interquartil, os valores que estivessem para alm dessas linhas, foram definidos e marcados pelo smbolo “+” como dados ultra discrepantes. Observou-se nos pacientes com DA um aumento da RP nas bandas delta e teta e uma diminuico da RP nas bandas alfa e beta, em comparaco com os sujeitos de controlo. Este efeito, chamado de “desaceleraco” mostrou conformidade com a literatura existente da DA [14, 46].

Pela observação dos valores de RP verificou-se que as diferenças mais significativas ($p < 0,0005$ e $AROC$ 0,86) foram registadas na faixa teta. Diferenças significativas também foram obtidas na banda beta1 ($p < 0,001$), banda beta2 ($p < 0,005$), na banda delta ($p < 0,01$) e banda alfa ($p < 0,05$). Os parâmetros que apresentaram os valores mais baixos de p foram conseguidos na banda teta e beta. Estes resultados revelaram consenso com alguns estudos como [52, 63, 80, 81, 82]. Segundo Poza [63], existem muitas diferenças entre pacientes com DA e sujeitos de controlo em $RP(teta)$ e $RP(beta)$. No presente estudo, isto também se verificou, mas a $RP(teta)$ superou a $RP(beta)$, o que não aconteceu no estudo de Poza [63].

Tabela 4.2 Médias e desvios padrão dos diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-Wallis, de parâmetros extraídos da PSD_n baseada na WT Bior 3.5, por indivíduo.

Parâmetros		Pacientes com DA	Controlos	
		Média \pm SD	Média \pm SD	p -valores
Potências Relativas	$RP(delta)$	0,15 \pm 0,05	0,12 \pm 0,01	0,00783
	$RP(teta)$	0,24 \pm 0,05	0,17 \pm 0,04	0,00041
	$RP(alfa)$	0,22 \pm 0,06	0,27 \pm 0,05	0,01576
	$RP(beta1)$	0,07 \pm 0,01	0,10 \pm 0,01	0,00053
	$RP(beta2)$	0,20 \pm 0,05	0,25 \pm 0,04	0,00459
	$RP(gama)$	0,11 \pm 0,04	0,11 \pm 0,03	N.S.
Rácios espectrais	r_1	0,88 \pm 0,59	1,80 \pm 0,64	0,00053
	r_2	1,75 \pm 0,68	2,55 \pm 0,59	0,00053
	r_3	2,00 \pm 0,98	2,84 \pm 0,49	0,00207
	r_4	1,45 \pm 0,77	2,07 \pm 0,42	0,00294

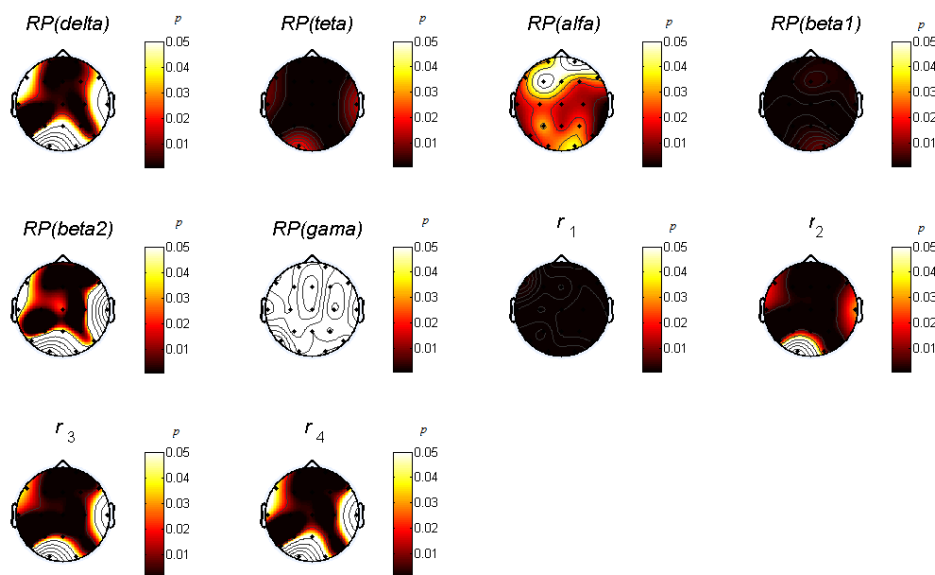
Tabela 4.3 Resultados de discriminação, obtidos através da curva ROC, dos parâmetros extraídos da PSD_n baseada na WT Bior 3.5, por indivíduo.

Parâmetros		Sensibilidade	Especificidade	Precisão	$AROC$
Potência Relativa	$RP(delta)$	92,90%	60,00%	73,50%	0,77
	$RP(teta)$	92,90%	75,00%	82,40%	0,86
	$RP(alfa)$	78,60%	65,00%	70,60%	0,75
	$RP(beta1)$	78,60%	75,00%	76,50%	0,85
	$RP(beta2)$	100,00%	65,00%	79,40%	0,79
	$RP(gama)$	50,10%	75,00%	61,80%	0,56
Rácios espectrais	r_1	92,90%	70,00%	79,40%	0,85
	r_2	85,70%	80,00%	82,40%	0,85
	r_3	92,90%	70,00%	79,40%	0,81
	r_4	71,40%	80,00%	76,50%	0,80

Pela observação da Tabela 4.2 e da Tabela 4.3 pode ver-se que os rácios espectrais apresentaram bons resultados ($p < 0,005$ com $r_3^{(WT)}$ e $r_4^{(WT)}$ e $p < 0,001$ com o $r_1^{(WT)}$ e $r_2^{(WT)}$, os últimos com um parâmetro $AROC$ óptimo (0,85)). Todavia, ao contrário do que aconteceu no estudo de Poza [63], os p -valores de $r_1^{(WT)}$ e $r_2^{(WT)}$ não foram suficientemente baixos para superar os resultados do parâmetro $RP(teta)$ (p -valor $0,00053 > 0,00041$). Além disso, nesse mesmo estudo MEG [63] obteve-se em r_3 e r_4 p -valores

mais baixos do que nas *RPs*. Porém, no presente estudo a $RP(beta1)^{(WT)}$ e a $RP(teta)^{(WT)}$ permitiram alcançar mais diferenças significativas entre os pacientes com DA e os sujeitos de controlo do que $r_3^{(WT)}$ e $r_4^{(WT)}$. Também se observou no presente estudo que os *p*-valores de $r_2^{(WT)}$, $r_3^{(WT)}$ e $r_4^{(WT)}$ foram muito baixos, o que levou a concluir que os pacientes se encontravam num estágio avançado da DA. Há que referir que se está a comparar um estudo MEG com um de EEG, o que implica que pequenas diferenças no sinal devem ser consideradas, porque apesar de os sinais terem a mesma origem foram medidos de maneira diferente e apresentaram uma frequência de amostragem diferente [83]. Deve atender-se ainda ao facto de que os processamentos aplicados foram ligeiramente diferentes no estudo MEG e no presente estudo. Fazendo uma análise mais aprofundada das diferenças entre os pacientes de DA e os sujeitos de controlo, neste caso por eléctrodo (Figura 4.8), verificou-se que os parâmetros que apresentaram mais diferenças nos 19 eléctrodos foram os r_1 , r_2 , $RP(teta)$, $RP(beta1)$, os mesmos que apresentaram maiores diferenças por indivíduo. Esta coincidência permitiu fazer uma pré-selecção das características que melhor discriminaram a DA em relação aos sujeitos de controlo.

Figura 4.8 Cabeças que representam as diferenças entre os pacientes de Alzheimer e os sujeitos de controlo por eléctrodo com um $p < 0.05$.



4.2.2.2 Parâmetros *RP* e rácios espectrais por eléctrodo

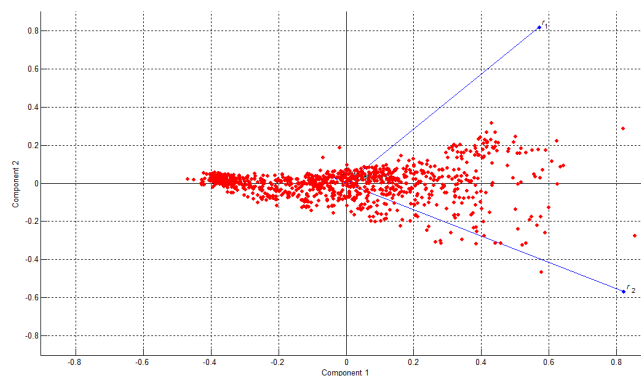
Devido à pequena dimensão da BD, o sinal de cada eléctrodo foi calculado pela média do sinal ao longo dos sucessivos segmentos de 5 segundos de duração - PSD_n por eléctrodo, constituindo um par

entrada/saída da RNA e da RL. Os eléctrodos pertencentes ao mesmo indivíduo foram forçados a estar no mesmo conjunto (treino ou de teste).

4.2.2.2.1 Classificação dos sinais EEG

Foram calculadas as componentes principais das 4 características (r_1 , r_2 , $RP(teta)$, $RP(beta1)$), obtidas pela PSD da WT Bior 3.5 que apresentaram os 4 p -valores mais baixos por indivíduo, pela técnica de PCA. De forma a poder verificar o modo como cada característica se comportava e contribuía para as componentes 1 e 2, foram projectados os vectores próprios das características num biplot (Figura 4.9).

Figura 4.9 Biplot das componentes principais das características r_1 , r_2 , $RP(beta1)$, $RP(teta)$.



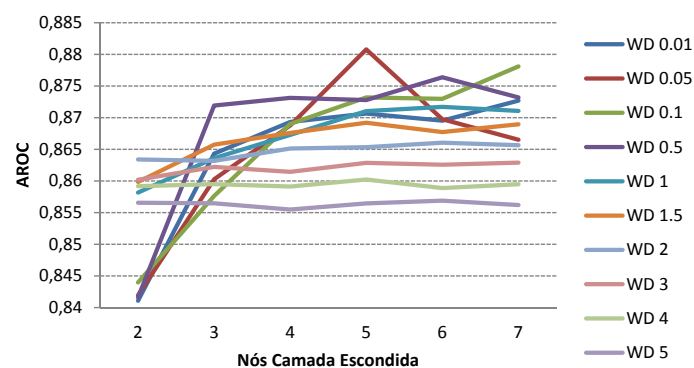
Verificou-se que a componente 2 fez a distinção entre a característica r_1 e a característica r_2 , pois exibiu um valor positivo e negativo, respectivamente, para as duas características. Os dois vectores representativos de r_1 e r_2 apresentaram um comprimento elevado, revelador de que eram as que mais contribuiriam para as componentes principais. Mostraram também uma direcção oposta, o que significa que as duas variáveis não apresentaram valores sobrepostos, constituindo assim uma boa combinação para a entrada da RNA e da RL. As outras características apresentaram um vector de pequena dimensão, pelo que contribuíram pouco para as componentes principais e poderiam induzir em erro a RNA e RL (diminuição do parâmetro $AROC$ e da precisão) ou pura e simplesmente, não contribuiriam significativamente para uma melhoria na discriminação entre os pacientes com DA e os sujeitos controlo, quando combinadas entre elas ou com os parâmetros r_1 e r_2 . Resumindo, a melhor combinação foi dada pelo par r_1 e r_2 pela análise da Figura 4.9. Assim, a RNA e a RL utilizadas apresentaram dois nós de entrada (Tabela 4.4).

Tabela 4.4 Conjuntos de Treino e de Teste.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	2
2x494	2x152	

Dividiram-se os 646 casos no conjunto de treino e de teste. O conjunto de treino ficou constituído por 494 e o de teste por 152 casos, como mostra a Tabela 4.4. A RNA apresentou uma função de activação Logsig, o SCG como algoritmo de treino e a entropia cruzada como função de erro. Para evitar o treino excessivo da RNA, foi calculado o parâmetro ideal de WD, que impediu os pesos de participarem plenamente no processo de modelagem da RNA para o conjunto de treino, evitando o “overfitting”. A fim de seleccionar o melhor parâmetro de WD e o número de nós da camada escondida para uma melhor classificação do conjunto de teste, usou-se o gráfico representado na Figura 4.10. Nesse gráfico pode ser visualizada a distribuição do parâmetro *AROC*, resultante da análise ROC dos resultados de classificação por validação cruzada da RNA deixando um de fora, do conjunto de treino, ao longo de um número variável de nós na Camada Escondida com diferentes parâmetros de WD. Este tipo de análise WD/nós, permitiu evitar a utilização do conjunto de validação e otimizar o número de nós da camada escondida.

Figura 4.10 Valores de *AROC* ao longo dos nós da camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Após uma breve análise do gráfico da Figura 4.10, escolheu-se 0.05 de WD e 5 nós na camada escondida para classificar o conjunto de teste pela RNA, pois estes parâmetros forneceram mais *AROC* pela RNA, no conjunto de treino, pelo processo de classificação por validação cruzada deixando um de fora. O conjunto de treino e de teste utilizados na metodologia para a RNA foram utilizados também para a RL. Finalmente, os resultados foram avaliados por alguns parâmetros extraídos da curva ROC (Tabela 4.5).

Tabela 4.5 Resultados da classificação do grupo de teste.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,97	0,80
Sensibilidade	92,1%	65,4%
Especificidade	90,8%	87,2%
Precisão	91,5%	74,9%

Para verificar o comportamento geral da RL e da RNA com 0.05 de WD e 5 nós na camada escondida, ao conjunto geral de 646 casos efectuou-se o processo de validação cruzada deixando um de fora pela RNA e pela RL, os resultados foram os apresentados na Tabela 4.6

Tabela 4.6 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,87	0,83
Sensibilidade	78,2%	70,8%
Especificidade	82,0%	79,3%
Precisão	79,7%	74,3%

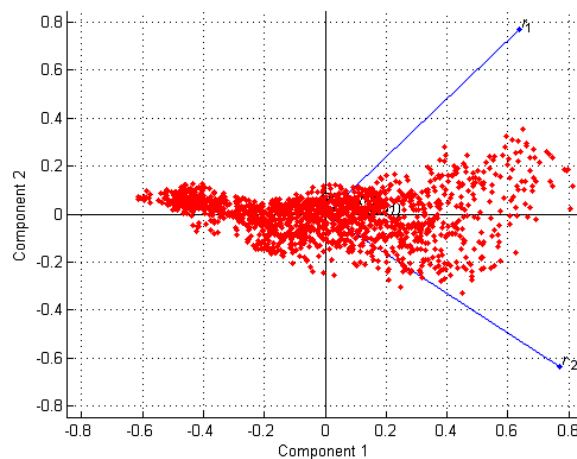
4.2.2.3 Parâmetros *RP* e rácios espectrais - “*grand average*” de 5s

Neste caso, cada período de 5 segundos do sinal que foi calculado pela média dos 19 eléctrodos por cada 5s, para produzir a chamada “*grand average*” de 5 segundos de duração, corresponde a um par de entrada/saída da RNA e RL . As sucessivas “*grand averages*” por indivíduo foram obrigadas a pertencer ao mesmo conjunto de treino ou de teste.

4.2.2.3.1 Classificação dos sinais EEG

Foram calculadas as componentes principais das 4 características (r_1 , r_2 , $RP(teta)$, $RP(beta1)$), obtidas pela *PSD* da WT Bior 3.5 que apresentaram os 4 *p*-valores mais baixos por indivíduo, pela técnica de PCA. De forma a poder verificar o modo como cada variável se comportava e contribuía para as componentes 1 e 2, foram projectados os vectores próprios das características num gráfico (biplot) xy, como mostra a Figura 4.11.

Figura 4.11 Biplot das componentes principais das características r_1 , r_2 , $RP(beta1)$, $RP(teta)$.



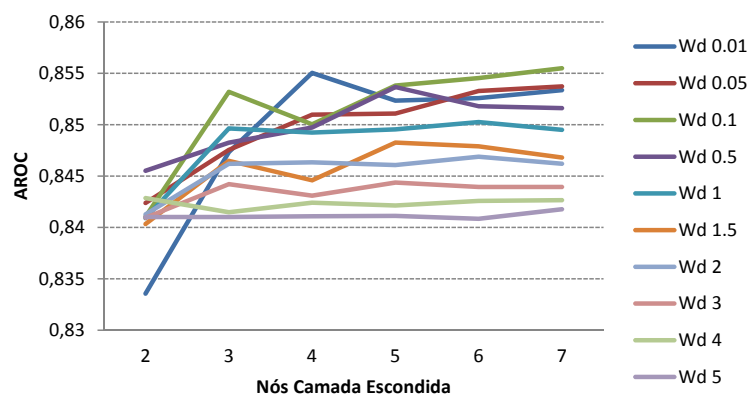
Verificou-se que a componente 2 fez a distinção entre a característica r_1 e a característica r_2 , pois mostrou um valor positivo e negativo, respectivamente, para as duas características. Os dois vectores representativos de r_1 e r_2 apresentaram um comprimento elevado, demonstrativo de que eram as que mais contribuiriam para as componentes principais. Evidenciaram também uma direcção oposta, o que significa que as duas variáveis não apresentaram valores sobrepostos, constituindo assim uma boa combinação para a entrada da RNA e da RL. As outras características revelaram um vector de pequena dimensão, pelo que contribuíram pouco para as componentes principais e poderiam induzir em erro a RNA e RL (diminuição do parâmetro *AROC* e da precisão), ou pura e simplesmente, não contribuiriam significativamente para uma melhoria na discriminação entre os pacientes com DA e os sujeitos de controlo, quando combinadas entre elas ou com os parâmetros r_1 e r_2 . Em suma, a melhor combinação foi dada pelo par r_1 e r_2 , e a RNA e a RL que se utilizaram apresentaram dois nós de entrada (Tabela 4.7).

Tabela 4.7 Conjuntos de Treino e de Teste.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	
2x780	2x287	2

Dividiram-se os 1067 casos em conjunto de treino e de teste. O conjunto de treino ficou constituído por 780 e o de teste por 287 casos (Tabela 4.7). Utilizou-se novamente uma RNA de reconhecimento de padrões com a função de activação Logsig, com o SCG como o algoritmo de treino e com a entropia cruzada como a função de erro. Para evitar o treino excessivo da RNA ao conjunto de treino, foi calculado o parâmetro ideal de WD. A fim de seleccionar o melhor valor de WD e o número de nós da camada escondida para uma melhor classificação do conjunto de teste, usou-se o gráfico da Figura 4.12.

Figura 4.12 Valores de *AROC* ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Depois de se analisar o gráfico da Figura 4.12, escolheu-se 0.1 de WD e 7 nós na camada escondida, pois estes parâmetros forneceram mais *AROC* pela RNA, para o conjunto de treino, no processo de classificação por validação cruzada deixando um de fora. O conjunto de treino e de teste utilizados para a RNA foram também utilizados para a RL. Finalmente, os resultados foram avaliados por alguns parâmetros extraídos da curva ROC (Tabela 4.8).

Tabela 4.8 Resultados da classificação do grupo de teste.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,90	0,80
Sensibilidade	87,0%	65,4%
Especificidade	88,0%	87,2%
Precisão	87,5%	74,9%

Para verificar o comportamento geral da RL e da RNA com 0.1 de WD e 7 nós na camada escondida, ao conjunto geral de 1067 casos procedeu-se ao processo de validação cruzada deixando um de fora, pela RNA e pela RL, e os resultados foram os constantes da Tabela 4.9.

Tabela 4.9 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,85	0,79
Sensibilidade	74,8%	70,0%
Especificidade	77,6%	72,8%
Precisão	76,1%	71,5%

4.2.3 Utilização de alguns parâmetros espectrais

Foram calculadas 4 características espectrais (*SE*, *SEF95%*, *MF*, *IAF*), conforme mostram as equações 2.22 à 2.25, de forma a poder resumir a dinâmica do sinal EEG a partir da PSD_n .

4.2.3.1 Análise estatística / selecção de características

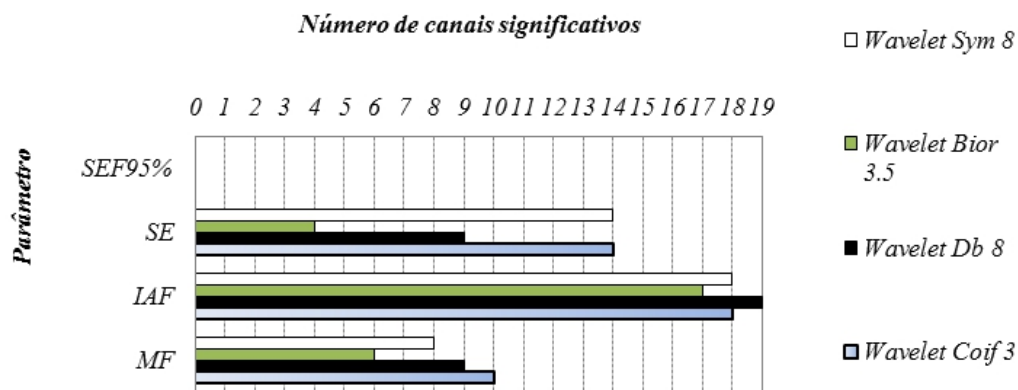
Procedeu-se à análise descritiva das distribuições de parâmetros obtidos da PSD_n por indivíduo (pacientes com DA e sujeitos de controlo). Analisou-se a normalidade das distribuições com o teste de KS e a homocedasticidade pelo teste de Levene. Estes parâmetros foram obtidos utilizando a WT Bior 3.5 escolhida de forma aleatória para o efeito.

Atendendo à Tabela 4.10 constatou-se que as distribuições de parâmetros não atenderam à hipótese dos testes paramétricos. Por isso, estudaram-se as diferenças entre os grupos por meio do teste de

Tabela 4.10 Teste de KS e teste de Levene das distribuições das RPs e dos rácios espectrais, obtidos por indivíduo

Parâmetros	Teste KS		Teste Levene	
	Teste Hipótese nula	<i>p</i> -valor	Teste Hipótese nula	<i>p</i> -valor
<i>SEF95%</i>	0	1	0	0,416
<i>SE</i>	1	0,003	1	0,001
<i>IAF</i>	1	0,007	0	0,335
<i>MF</i>	1	0,01	0	0,107

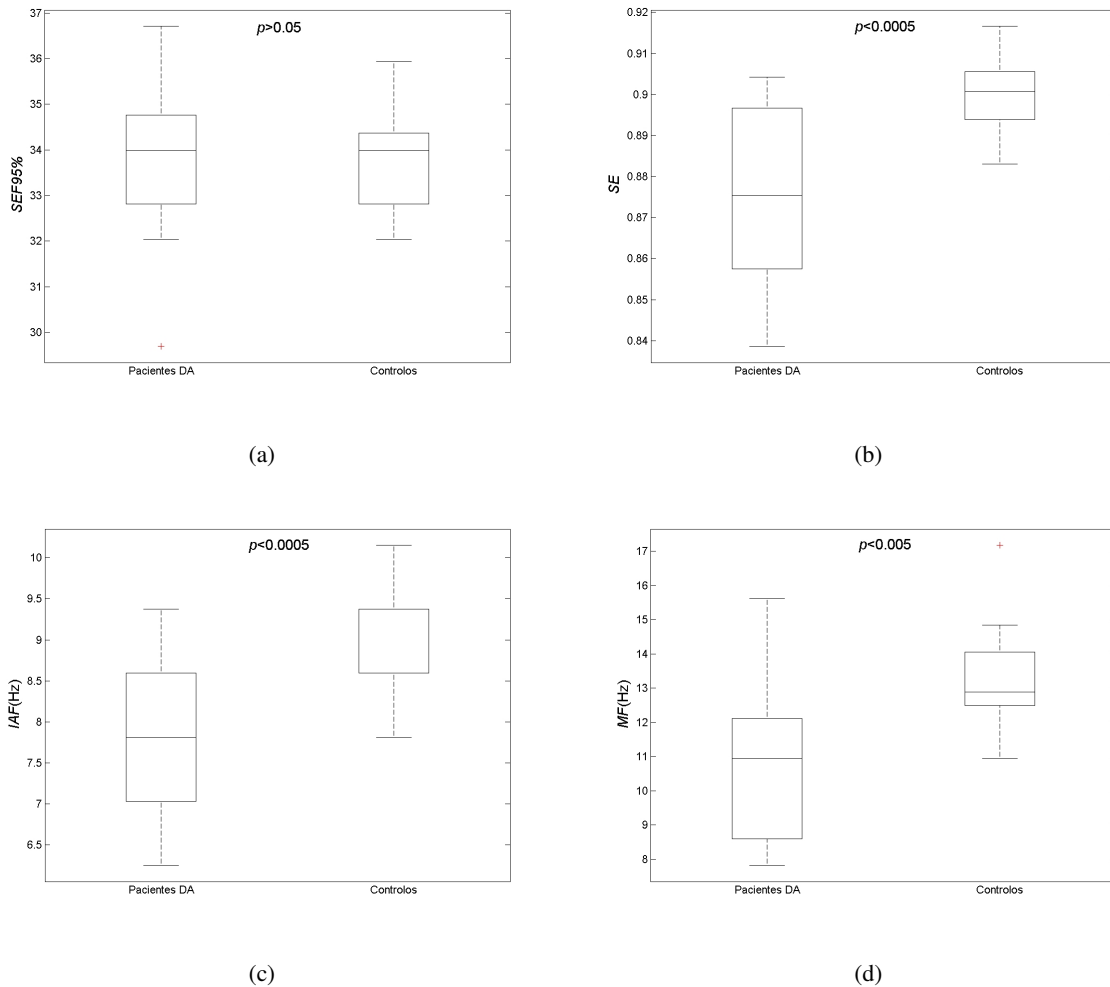
Kruskal-Wallis que é um teste não-paramétrico. De forma a seleccionar a Wavelet que se ajustou melhor à dinâmica do sinal de EEG nesta abordagem, procedeu-se à comparação entre os resultados dos parâmetros dos pacientes com DA e dos sujeitos de controlo extraídos da PSD_n por eléctrodo para um conjunto de *WTs*, pelo teste de Kruskal-Wallis com a correcção de Bonferroni para $p < 0.05$. O gráfico que permitiu a selecção da Wavelet encontra-se na Figura 4.13.

Figura 4.13 Número de canais significativos para os processamentos envolvidos neste estudo com *p*-valores < 0.05 obtidos através do teste de Kruskal-Wallis com correcção de Bonferroni.

Através da Figura 4.13 foi perceptível que não existe uma Wavelet que se superiorize muito em relação às outras na discriminação da doença de DA. Contudo, utilizou-se a Wavelet que teve uma boa performance na discriminação da DA. A selecção recaiu sobre a Wavelet Coif.3, e a partir daí, foi utilizada para classificar os sinais de EEG nesta abordagem. No sentido de verificar as distribuições dos parâmetros obtidos pela PSD_n por indivíduo, utilizaram-se novamente os diagramas de caixa, para de uma forma mais resumida, visual e intuitiva se poder verificar as diferenças entre as distribuições de parâmetros obtidas para os pacientes com DA e para os sujeitos de controlo.

Pelos diagramas de caixa da Figura 4.14, observou-se que os sujeitos de controlo apresentaram uma média de valores superior relativamente aos pacientes com DA em todos os parâmetros, excepto em *SEF95%*. Estas tendências já se tinham notado em trabalhos anteriores para sinais de MEG e para EEG [62, 84], excepto no caso do parâmetro *SEF95%*, pois neste caso este não conseguiu fazer uma boa

Figura 4.14 Diagramas de caixa que mostram a distribuição dos valores de alguns parâmetros espectrais pela WT COIF 3.



discriminação das distribuições. Os parâmetros que exibiram os p -valores mais baixos foram o IAF e a SE . Observando ainda o diagrama de caixa relativo à SE , pode ver-se que os pacientes com DA em geral, apresentaram valores de SE mais baixos do que os sujeitos de controlo, permitindo dizer que os sinais de EEG de pacientes com DA são mais regulares do que os sinais dos sujeitos de controlo, concordando com o estudo de Poza [62] e o de Abásolo [84]. A partir da observação das Tabela 4.11 e a Tabela 4.12, viu-se que os parâmetros SE e IAF permitiram obter um p -valor igual a 0.0003. Por isso, os dois parâmetros discriminam bem as diferenças entre os dois grupos envolvidos no estudo porque o p -valor foi muito baixo. A MF também possibilitou uma boa discriminação, ao contrário do $SEF95\%$ que apresentou o p -valor > 0.05 , o que equivale a dizer que este parâmetro não apresentou diferenças significativas na distribuição de valores. Procedeu-se a uma análise mais promenorizada das diferenças entre os pacientes de DA e os sujeitos de controlo, neste caso por eléctrodo (Figura 4.15) e reparou-se que os parâmetros que apresentaram mais diferenças nos 19 eléctrodos foram os SE , IAF e MF . Facultaram desse modo a

realização de uma pré selecção das características que melhor discriminam a DA em relação aos sujeitos de controlo.

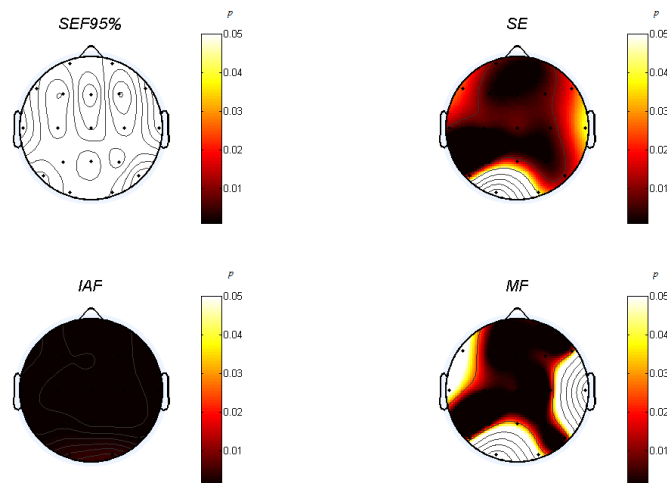
Tabela 4.11 Médias e desvios padrão dos Diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-Wallis, de parâmetros espectrais extraídos da PSD_n baseada na WT Coif. 3, por indivíduo.

Parâmetros	Pacientes com DA	Controlos	p -valores
	Média \pm SD	Média \pm SD	
$SEF95\%$	34,38 \pm 1,54	34,38 \pm 1,1	0,9573
SE	0,89 \pm 0,02	0,90 \pm 0,01	0,0003
IAF	7,81 \pm 1,01	8,59 \pm 0,79	0,0003
MF	11,72 \pm 2,60	12,50 \pm 1,8	0,0041

Tabela 4.12 Resultados de discriminação obtidos através da curva ROC, dos parâmetros espectrais extraídos da PSD_n baseada na WT Coif. 3, por indivíduo.

Parâmetros	Sensibilidade (%)	Especificidade (%)	Precisão (%)	$AROC$
$SEF95\%$	100,00%	0,00%	60,82%	0,51
SE	95,00%	64,29%	82,35%	0,87
IAF	70,00%	85,71%	76,47%	0,86
MF	75,00%	78,57%	76,47%	0,79

Figura 4.15 Cabeças que representam as diferenças entre os pacientes de Alzheimer e os sujeitos de controlo por eléctrodo com um $p < 0.05$.



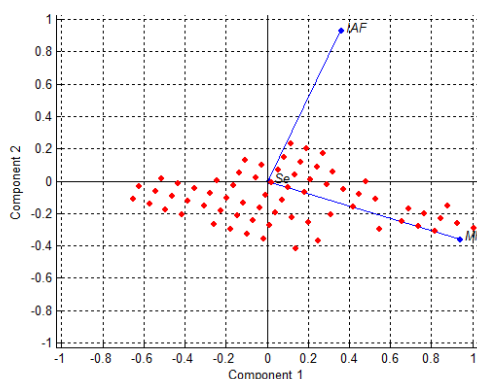
4.2.3.2 Parâmetros espectrais por eléctrodo

Devido à pequena dimensão da BD, nesta abordagem, o sinal de cada eléctrodo foi calculado pela média do sinal ao longo dos sucessivos segmentos de 5 segundos de duração - PSD_n por eléctrodo, constituindo um par entrada/saída da RNA e da RL. Os eléctrodos pertencentes ao mesmo indivíduo foram forçados a estar no mesmo conjunto de treino ou de teste.

4.2.3.2.1 Classificação dos sinais EEG

Foram calculadas as duas componentes principais das 3 características (SE , IAF , MF) que apresentaram o p -valor mais baixo por eléctrodo, pela técnica de PCA. De forma a poder comprovar o modo como cada variável se comportava e contribuiu para as componentes 1 e 2, foram projectados os vectores próprios das características num gráfico conforme mostra a figura 4.16.

Figura 4.16 Biplot das componentes principais das características SE , IAF , MF .



Verificou-se que a componente principal 2 fez a distinção entre a característica MF e a característica IAF , pois apresentou um valor positivo e negativo, respectivamente, para as duas características. Os dois vectores representativos de MF e IAF apresentaram um comprimento elevado, o que significa que eram as que mais contribuíram para as componentes principais. Mostraram também uma direcção oposta, e por isso não apresentaram valores sobrepostos, constituindo assim uma boa combinação para a entrada da RNA e da RL. SE exibiu um vector de pequena dimensão e contribuiu pouco para as componentes principais podendo induzir em erro a RNA e a RL (diminuição do parâmetro $AROC$ e da precisão) ou simplesmente, não contribuiria significativamente para uma melhoria na discriminação entre a DA e os sujeitos de controlo, quando combinadas com os parâmetros MF e IAF . Concluindo, a melhor combinação foi dada pelo par MF e IAF (gráfico da Figura 4.16) e assim a RNA e a RL apresentam dois nós de entrada (Tabela 4.13).

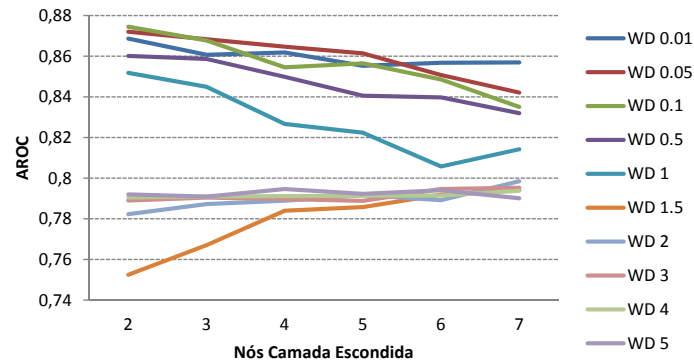
Tabela 4.13 Conjuntos de Treino e de Teste.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	2
2x494	2x152	

Dividiram-se os 646 casos no conjunto de treino e de teste. O conjunto de treino ficou constituído por 494 e o de teste por 152 casos (Tabela 4.13). A RNA usada continha a função de activação Logsig, o SCG como algoritmo de treino e a entropia cruzada como a função de erro. Para evitar o treino excessivo

da RNA, foi calculado o parâmetro ideal de WD. Para se seleccionarem os melhores parâmetros de WD e o número de nós da camada escondida, usou-se o gráfico da Figura 4.17.

Figura 4.17 Valores de *AROC* ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Feita uma breve análise do gráfico, escolheu-se 0.1 de WD e 2 nós na camada escondida, pois estes parâmetros forneceram mais *AROC* pela RNA, para o conjunto de treino, no processo de validação cruzada deixando um de fora. O conjunto de treino e de teste utilizados para a RNA também foram usados para a RL. Finalmente, os resultados foram avaliados por alguns parâmetros extraídos da curva ROC.

Tabela 4.14 Resultados da classificação do grupo de teste.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,93	0,77
Sensibilidade	77,6%	66,4%
Especificidade	92,1%	82,2%
Precisão	84,9%	72,9%

Para verificar o comportamento geral da RL e da RNA com 0.1 de WD e 2 nós na camada escondida ao conjunto geral de 646 casos, procedeu-se ao processo de classificação por validação cruzada deixando um de fora pela RNA e pela RL, e os resultados foram os registados na Tabela 4.15.

Tabela 4.15 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,87	0,78
Sensibilidade	75,5%	71,0%
Especificidade	82,0%	69,8%
Precisão	78,2%	71,2%

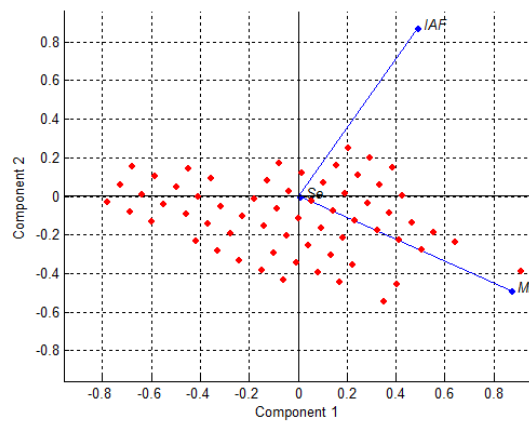
4.2.3.3 Parâmetros espectrais - “grand average” de 5s

Neste caso, cada período de 5 segundos do sinal que foi calculado pela média dos 19 eléctrodos por cada 5s para produzir a chamada “grand average” de 5 segundos de duração, corresponde a um par de entrada/saída da RNA e da RL. As sucessivas “grand averages” de 5s, por indivíduo, foram obrigadas a pertencer ao mesmo conjunto de treino ou de teste.

4.2.3.3.1 Classificação dos sinais EEG

Foram calculadas as componentes principais das 3 características (SE , IAF , MF) que apresentaram o p -valor mais baixo quer por indivíduo, quer por eléctrodo, pela técnica de PCA. De forma a poder verificar o modo como cada variável se comportava e contribuiu para as componentes 1 e 2, foram projectados os vectores próprios das características num gráfico xy, conforme consta na Figura 4.18.

Figura 4.18 Biplot das componentes principais das características SE , IAF , MF .

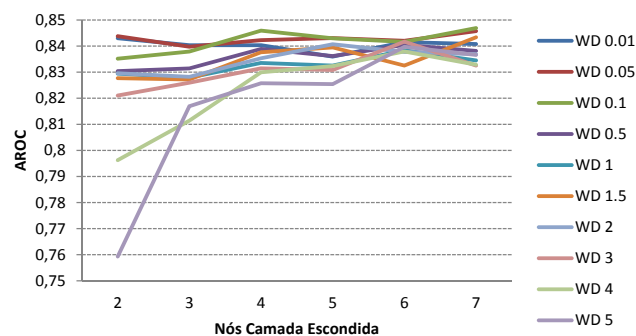


Constatou-se que a componente principal 2 fez a distinção entre a característica MF e a característica IAF , pois apresentou um valor positivo e negativo, respectivamente, para as duas características. Os dois vectores representativos de MF e IAF mostraram um comprimento elevado, o que significa que eram as que mais contribuíram para as componentes principais. Evidenciaram também uma direcção oposta, e por isso não apresentaram valores sobrepostos, constituindo assim uma boa combinação para a entrada da RNA e RL. Dado que a SE exibiu um vector de pequena dimensão contribuiu pouco para as componentes principais e poderia induzir em erro a RNA e a RL (diminuição do parâmetro $AROC$ e da precisão) ou simplesmente, não contribuiria significativamente para uma melhoria na discriminação entre pacientes com DA e os sujeitos de controlo, quando combinadas com os parâmetros MF e IAF . Concluindo, a melhor combinação foi dada pelo par MF e IAF (gráfico da Figura 4.18) e assim a RNA e a RL apresentam dois nós de entrada (Tabela 4.16).

Tabela 4.16 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	2
2x780	2x287	

Dividiram-se os 1067 casos em conjunto de treino e de teste. O conjunto de treino ficou constituído por 780 e o de teste por 287 casos (Tabela 4.16). A RNA usada continha uma função de activação Logsig, o SCG como o algoritmo de treino e a entropia cruzada como a função de erro. Para evitar o treino excessivo da RNA, foi calculado o parâmetro ideal de WD. A fim de seleccionar o melhor WD e o número de nós da camada escondida para uma melhor classificação do conjunto de teste, usou-se o gráfico presente na Figura 4.19. Após uma análise desse gráfico, escolheu-se 0.1 de WD e 7 nós na camada escondida para classificar o conjunto de teste pela RNA, pois estes parâmetros forneceram mais AROC pela RNA, para o conjunto de treino, no processo de validação cruzada deixando um de fora. O conjunto de treino e de teste utilizados para a RNA foram igualmente usados para a RL.

Figura 4.19 Valores de AROC ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.

Os resultados foram avaliados por alguns parâmetros extraídos da curva ROC (Tabela 4.17).

Tabela 4.17 Resultados da classificação do grupo de teste.

Parâmetros	Método	
	RNA	RL
AROC	0,9	0,76
Sensibilidade	88,0%	70,4%
Especificidade	87,0%	82,3%
Precisão	87,5%	76,3%

Para verificar o comportamento geral da RL e da RNA com 0.1 de WD e 7 nós na camada escondida, ao conjunto geral de 1067 casos procedeu-se ao processo de classificação por validação cruzada deixando um de fora pela RNA e pela RL e os resultados foram os mostrados na Tabela 4.18.

Tabela 4.18 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,85	0,8
Sensibilidade	77,6%	71,1%
Especificidade	74,8%	73,9%
Precisão	76,1%	72,6%

4.3 Classificação dos sinais EEG utilizando a PSD baseada na STFT

A técnica de *PSD* obtida através da STFT foi utilizada para a extracção de características dos sinais EEG. As abordagens aqui aplicadas foram as mesmas que se utilizaram na classificação dos sinais EEG utilizando a *PSD* baseada na WT.

4.3.1 Processamento do sinal EEG e extracção de características

Foi calculada a *PSD* baseada na STFT. Foi determinada a média das várias *PSDs* ao longo dos sucessivos segmentos de 5s de duração do sinal por eléctrodo (PSD_n por eléctrodo), e de seguida a média das várias PSD_n por eléctrodo (PSD_n por indivíduo) (ver exemplo, Figura 4.4).

4.3.2 Utilização dos parâmetros *RP* e rácios espectrais

Foram calculadas 10 características (*RP* em delta, teta, alfa, beta1, beta2, gama, bem como as relações espectrais associadas r_1, r_2, r_3, r_4), de acordo com as equações 2.26 à 2.35, de forma a poder resumir a dinâmica do sinal EEG a partir da PSD_n .

4.3.2.1 Análise estatística / selecção de características

Foi realizada a análise descritiva da distribuição de parâmetros extraídos, dos EEG dos pacientes com DA e sujeitos de controlo. Analisou-se a normalidade da distribuição com o teste de KS e a homocedasticidade pelo teste de Levene (Tabela 4.19). No caso do teste de KS a distribuição de parâmetros em geral rejeitou a hipótese nula (normalidade da distribuição dos parâmetros) e no teste de Levene a distribuição de parâmetros de um modo geral aceitou a hipótese nula (homocedasticidade da distribuição dos parâmetros). Por conseguinte, concluiu-se que a distribuição não atendeu à hipótese dos testes paramétricos, e por isso, estudaram-se as diferenças entre os grupos por meio do teste de Kruskal-Wallis que é um teste não-paramétrico. A fim de se verificarem as distribuições dos parâmetros, obtidos pela PSD_n por indivíduo, dos pacientes de DA e dos sujeitos de controlo, utilizaram-se os diagramas de caixa (ver Figura 4.20 e Figura 4.21).

Tabela 4.19 Teste de Ks e teste de Levene das distribuições das RPs e dos rácios espectrais - por indivíduo

Parâmetros	Teste KS		Teste Levene	
	Teste Hipótese nula	p -valor	Teste Hipótese nula	p -valor
$RP(\delta)$	1	0,0488	1	0,0008
$RP(\theta)$	1	0,0004	1	0,0295
$RP(\alpha)$	1	0,0488	0	0,1454
$RP(\beta_1)$	1	0,0004	0	0,2582
$RP(\beta_2)$	1	0,0043	0	0,5131
$RP(\gamma)$	0	0,7596	0	0,7662
r_1	1	0,0078	0	0,8247
r_2	1	0,0027	0	0,3963
r_3	1	0,0043	0	0,0736
r_4	1	0,0119	0	0,1613

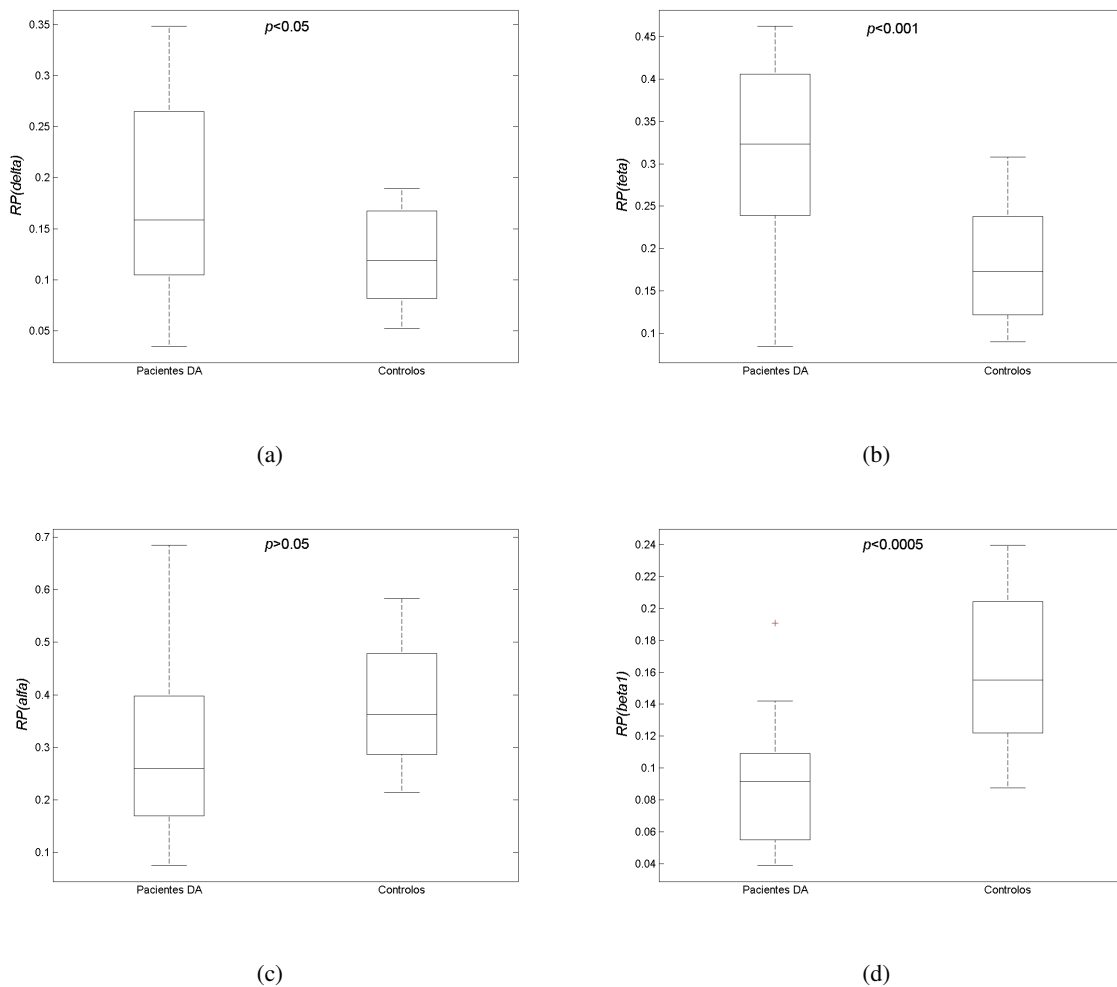
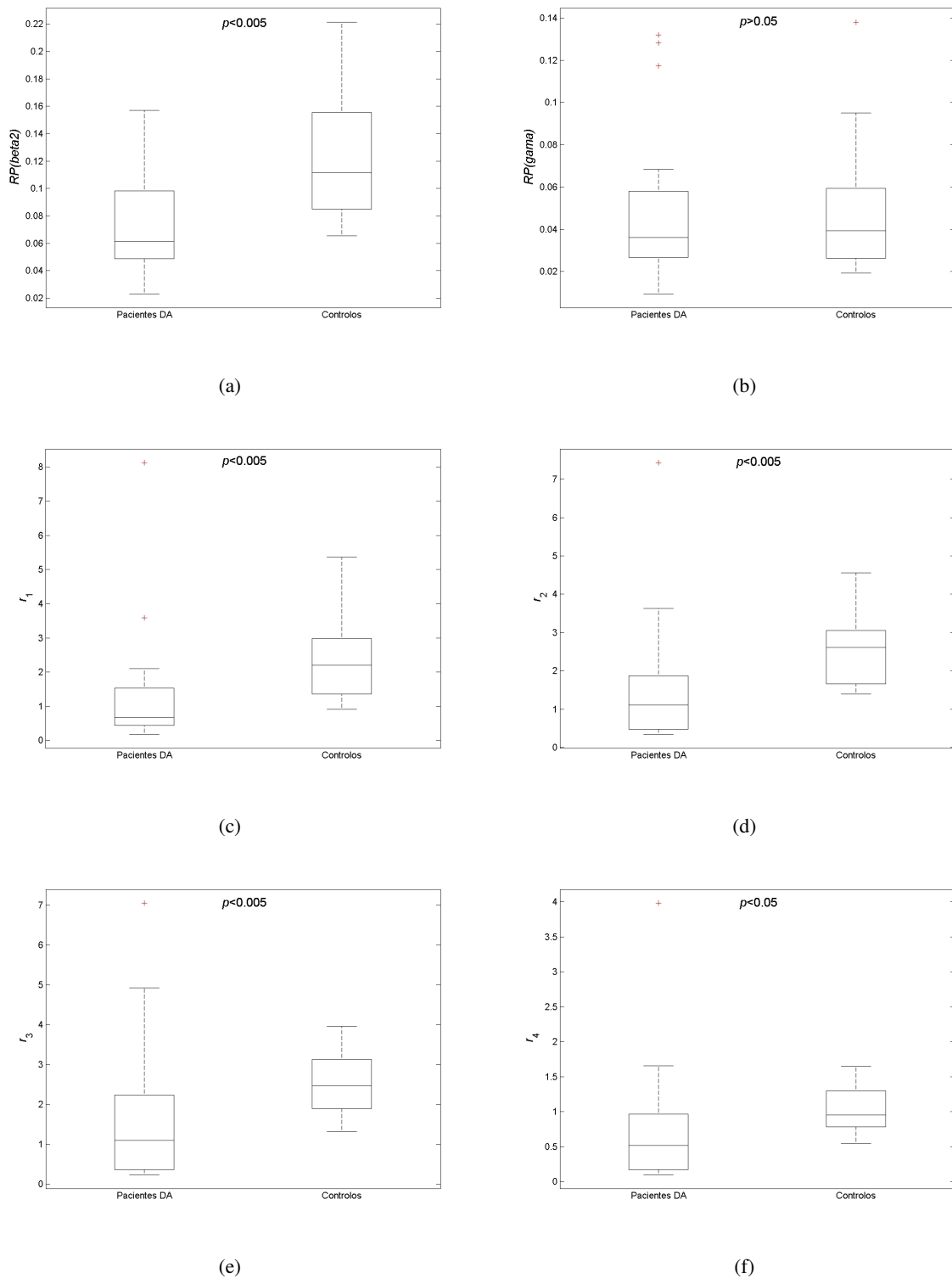
Figura 4.20 Diagramas de caixa que mostram a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais por indivíduo, obtidos pela PSD STFT.

Figura 4.21 Diagramas de caixa que mostram a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais por indivíduo, obtidos pela PSD STFT - continuação.



Pela observação dos Diagramas de caixa da Figura 4.20 e da Figura 4.21 nota-se nos pacientes com DA um aumento da RP nas faixas delta e teta e uma diminuição da RP nas faixas alfa e beta, em com-

paração com os sujeitos de controlo. Pela análise dos valores de RP pode verificar-se que as diferenças mais significativas ($p < 0,0005$ e 0.89 de $AROC$) foram obtidas na banda beta1. Diferenças significativas também foram alcançadas na banda teta ($p < 0,001$) e na banda beta2 ($p < 0,005$). Os parâmetros que apresentaram os valores mais baixos de p foram conseguidos na banda teta e beta. Estes resultados estão em consonância com alguns estudos [52, 63, 80, 81, 82].

Tabela 4.20 Médias e desvios padrão dos Diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-Wallis, de parâmetros extraídos da PSD_n baseada na STFT, por indivíduo.

Parâmetros	Pacientes com DA		Controlos	
		Média \pm SD	Média \pm SD	p -valores
Potências Relativas	$RP(\delta)$	0,15 \pm 0,1	0,12 \pm 0,04	NS
	$RP(\theta)$	0,32 \pm 0,1	0,17 \pm 0,04	0,00053
	$RP(\alpha)$	0,26 \pm 0,16	0,36 \pm 0,1	NS
	$RP(\beta_1)$	0,09 \pm 0,04	0,15 \pm 0,04	0,00012
	$RP(\beta_2)$	0,06 \pm 0,04	0,11 \pm 0,04	0,00411
	$RP(\gamma)$	0,036 \pm 0,04	0,039 \pm 0,033	N.S.
Rácios espectrais	r_1	0,67 \pm 1,79	2,2 \pm 1,24	0,00294
	r_2	1,11 \pm 1,64	2,61 \pm 0,93	0,00184
	r_3	1,1 \pm 1,76	2,47 \pm 0,81	0,00570
	r_4	0,52 \pm 0,89	0,9498 \pm 0,36	0,01576

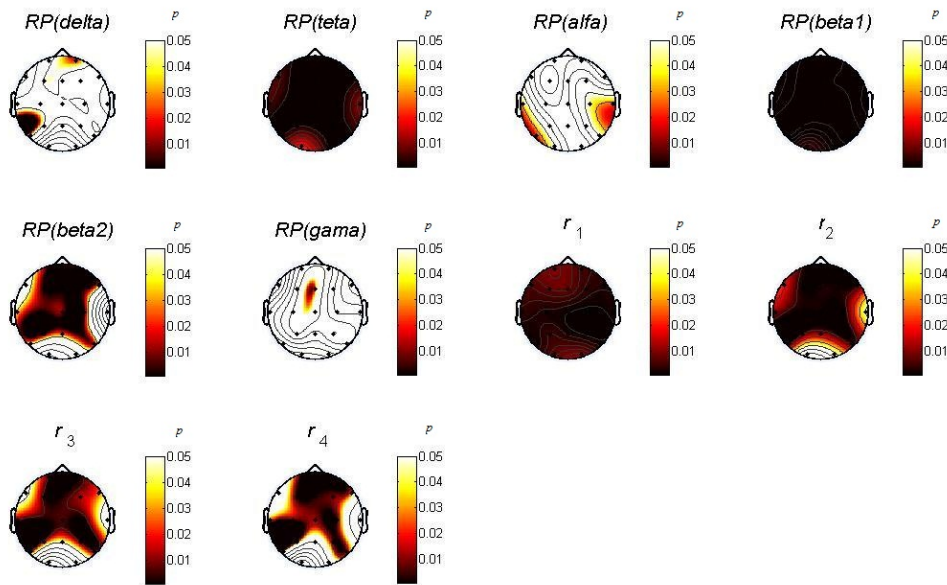
Tabela 4.21 Resultados de discriminação, obtidos através da curva ROC, dos parâmetros extraídos da PSD_n baseada STFT, por indivíduo.

Parâmetros	Sensibilidade	Especificidade	Precisão	$AROC$	
Potência Relativa	$RP(\delta)$	45%	100,00%	67,65%	0,66
	$RP(\theta)$	75,00%	92,86%	82,35%	0,85
	$RP(\alpha)$	70,00%	64,29%	67,65%	0,70
	$RP(\beta_1)$	80,00%	85,71%	82,35%	0,89
	$RP(\beta_2)$	70,00%	85,71%	76,47%	0,79
	$RP(\gamma)$	65,00%	57,14%	61,76%	0,56
Rácios espectrais	r_1	80,00%	71,43%	76,47%	0,80
	r_2	80,00%	71,43%	76,47%	0,82
	r_3	65,00%	92,86%	76,47%	0,78
	r_4	65,00%	85,71%	73,53%	0,75

Pela observação das Tabelas 4.20 e 4.21 pode ver-se que os rácios espectrais mostraram bons resultados ($p < 0,01$ com $r_3^{(STFT)}$ e $p < 0,005$ com o $r_1^{(STFT)}$ e $r_2^{(STFT)}$, os últimos com um parâmetro $AROC$ óptimo superior a 0.8). Neste estudo, a $RP(\beta_1)^{(STFT)}$ e a $RP(\theta)^{(STFT)}$ permitiram obter mais diferenças significativas entre os pacientes com DA e sujeitos de controlo do que $r_3^{(STFT)}$ e $r_4^{(STFT)}$. Procedendo a uma análise mais detalhada das diferenças entre os pacientes com DA e de sujeitos de controlo, neste caso por eléctrodo (Figura 4.22), verificou-se que os parâmetros que apresentaram mais diferenças nos 19 eléctrodos foram os r_1 , r_2 , $RP(\theta)$, $RP(\beta_1)$, os mesmos que apresentavam maiores diferenças

por indivíduo. O que possibilitou uma pré-selecção das características que discriminariam melhor a DA em relação aos sujeitos de controlo.

Figura 4.22 Cabeças que representam as diferenças entre os pacientes de DA e os sujeitos de controlo por eléctrodo com um $p < 0.05$.



4.3.2.2 Parâmetros RP e rácios espectrais por eléctrodo

Devido à pequena dimensão da BD, na abordagem empregue, o sinal de cada eléctrodo foi calculado pela média do sinal ao longo dos sucessivos segmentos de 5 segundos de duração - PSD_n por eléctrodo, constituindo um par entrada/saída da RNA e da RL. Os eléctrodos pertencentes ao mesmo indivíduo foram forçados a estar no mesmo conjunto (treino ou de teste).

4.3.2.2.1 Classificação dos sinais EEG

Foram calculadas as componentes principais das 4 características (r_1 , r_2 , $RP(teta)$, $RP(beta1)$), que apresentaram o p -valor mais baixo por eléctrodo pela técnica de PCA. De forma a poder verificar o modo como cada variável se comportava e contribuía para as componentes 1 e 2 foram projectados os vectores próprios das características num gráfico xy, que pode ser observado na figura 4.23. Os dois vectores que representam a $RP(beta1)$ e $RP(teta)$ mostraram uma direcção oposta, significa isto que os dois parâmetros não exibiram valores sobrepostos. Isto, e as grandes diferenças significativas entre os p -valores de $RP(beta1)$ e $RP(teta)$ e os p -valores de r_1 e r_2 (Confrontar Figura 4.23) indicaram que a $RP(beta1)$ e $RP(teta)$, constituiriam uma boa combinação para a entrada RNA e da RL. Em

suma, a melhor combinação foi dada pelo par $RP(beta1)$ e $RP(teta)$. A RNA e a RL que se utilizaram apresentaram dois nós de entrada (Tabela 4.22).

Figura 4.23 Biplot das componentes principais das características $r_1, r_2, RP(beta1), RP(teta)$.

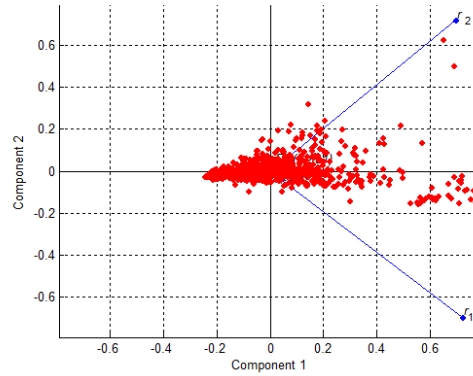
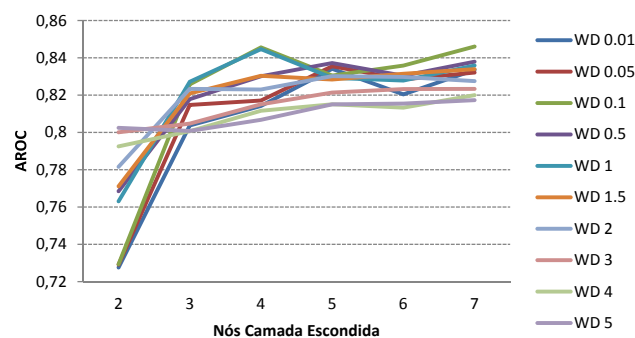


Tabela 4.22 Conjuntos de Treino e de Teste.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	
2x494	2x152	2

Dividiram-se os 646 casos no conjunto de treino e de teste. O conjunto de treino ficou constituído por 494 e o de teste por 152 casos (Tabela 4.22). A RNA usada continha a função de activação Logsig, o SCG como algoritmo de treino e a entropia cruzada como a função de erro. Procurando evitar o treino excessivo da RNA, calculou-se o parâmetro ideal de WD. Para se seleccionar o melhor WD e o número de nós da camada escondida, usou-se o gráfico da Figura 4.24.

Figura 4.24 Valores de AROC ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Depois de se analisar o gráfico da Figura 4.24, escolheu-se 1 de WD e 4 nós na camada escondida, pois estes parâmetros forneceram mais AROC pela RNA, para o conjunto de treino, no processo de classificação por validação cruzada deixando de um fora. O conjunto de treino e de teste usados para a RNA foram utilizados também para a RL. Finalmente, os resultados foram avaliados por alguns parâmetros extraídos da curva ROC (ver Tabela 4.23).

Tabela 4.23 Resultados da classificação do grupo de teste

Parâmetros	Método	
	RNA	RL
AROC	0,83	0,68
Sensibilidade	76,3%	71,2%
Especificidade	92,1%	78,2%
Precisão	84,6%	76,9%

Para verificar o comportamento geral da RL e da RNA com 1 de WD e 4 nós na camada escondida, ao conjunto geral de 646 casos procedeu-se ao processo de validação cruzada deixando um de fora pela RNA e pela RL, e os resultados foram os apresentados na Tabela 4.24.

Tabela 4.24 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
AROC	0,83	0,78
Sensibilidade	69,9%	65,0%
Especificidade	75,5%	77,8%
Precisão	73,2%	70,3%

4.3.2.3 Parâmetros RP e rácios espectrais - “grand average” de 5s

Neste caso, cada período de 5 segundos do sinal que foi calculado pela média dos 19 eléctrodos por cada 5s, para produzir a chamada “grand average” de 5 segundos de duração, corresponde a um par de entrada/saída da RNA e RL. As sucessivas “grand averages” 5s por indivíduo foram obrigadas a pertencer ao mesmo conjunto de treino ou de teste.

4.3.2.3.1 Classificação dos sinais EEG

Foram calculadas as duas componentes principais das 4 características (r_1 , r_2 , $RP(teta)$, $RP(beta1)$) que apresentaram o p -valor mais baixo quer por indivíduo, quer por eléctrodo, pela técnica de PCA. De modo a poder verificar o modo como cada variável se comportava e contribuía para as componentes principais 1 e 2, foram projectados os vectores próprios das características num gráfico que pode ser observado na Figura 4.25.

Os dois vectores que representam a $RP(beta1)$ e $RP(teta)$ mostraram uma direcção oposta, e por isso, não exibiram valores sobrepostos. Associando a isso, as diferenças significativas registadas entre os p -valores de $RP(beta1)$ e $RP(teta)$ e os p -valores de r_1 e r_2 (Confrontar Tabela 4.19), permitiram concluir que a $RP(beta1)$ e $RP(teta)$, constituiriam uma boa combinação para a entrada RNA e a RL. Concluindo, a melhor combinação foi dada pelo par $RP(beta1)$ e $RP(teta)$ depois de analisada a Figura 4.25, e a RNA / RL utilizadas apresentaram dois nós de entrada (Tabela 4.25).

Figura 4.25 Biplot das componentes principais das características r_1 , r_2 , $RP(beta1)$, $RP(teta)$.

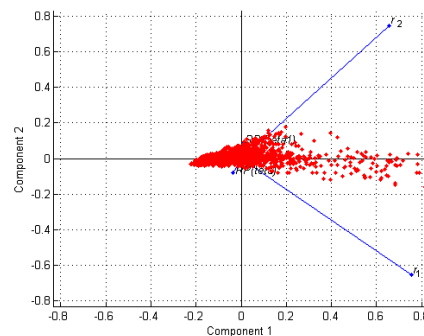
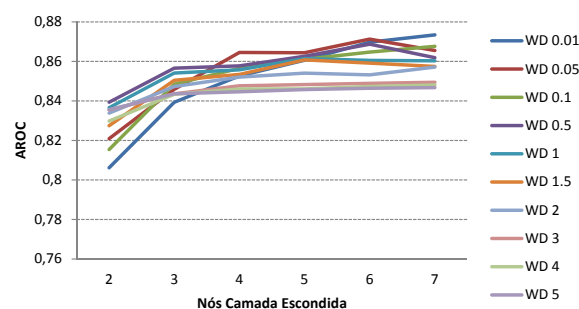


Tabela 4.25 Conjuntos de Treino e de Teste.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	
2x780	2x287	2

Dividiram-se os 1067 casos em conjunto de treino e de teste. O conjunto de treino ficou constituído por 780 e o de teste por 287 casos (Tabela 4.25). Utilizou-se novamente uma RNA de reconhecimento de padrões com a função de activação Logsig, com o SCG como o algoritmo de treino e com a entropia cruzada como a função de erro. Procurando seleccionar o melhor valor para WD e o número de nós da camada escondida para uma melhor classificação do conjunto de teste, usou-se o gráfico da Figura 4.26.

Figura 4.26 Valores de AROC ao longo dos nós da camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Após uma breve análise dos dados do gráfico da Figura 4.26, escolheu-se 0.01 de WD e 7 nós na camada escondida, dado que estes parâmetros forneceram mais *AROC* pela RNA. O conjunto de treino e de teste usados para a RNA foram também utilizados para a RL. Por último fez-se a avaliação dos resultados através de alguns parâmetros extraídos da curva ROC, constantes na Tabela 4.26.

Tabela 4.26 Resultados da classificação do grupo de teste.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,81	0,76
Sensibilidade	73,6%	71,2%
Especificidade	90,7%	67,8%
Precisão	83,3%	68,9%

Para verificar o comportamento geral da RL e da RNA com 0.01 de WD e 7 nós na camada escondida, ao conjunto geral de 1067 casos procedeu-se ao processo de validação cruzada deixando um de fora pela RNA e pela RL. Os resultados podem ser observados na Tabela 4.27.

Tabela 4.27 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,83	0,72
Sensibilidade	74,0%	63,3%
Especificidade	80,4%	75,9%
Precisão	76,8%	68,7%

4.3.3 Utilização de alguns parâmetros espectrais

Foram calculadas 4 características espectrais (*SE*, *SEF95%*, *MF*, *IAF*) de forma a poder resumir a dinâmica do sinal EEG a partir da PSD_n .

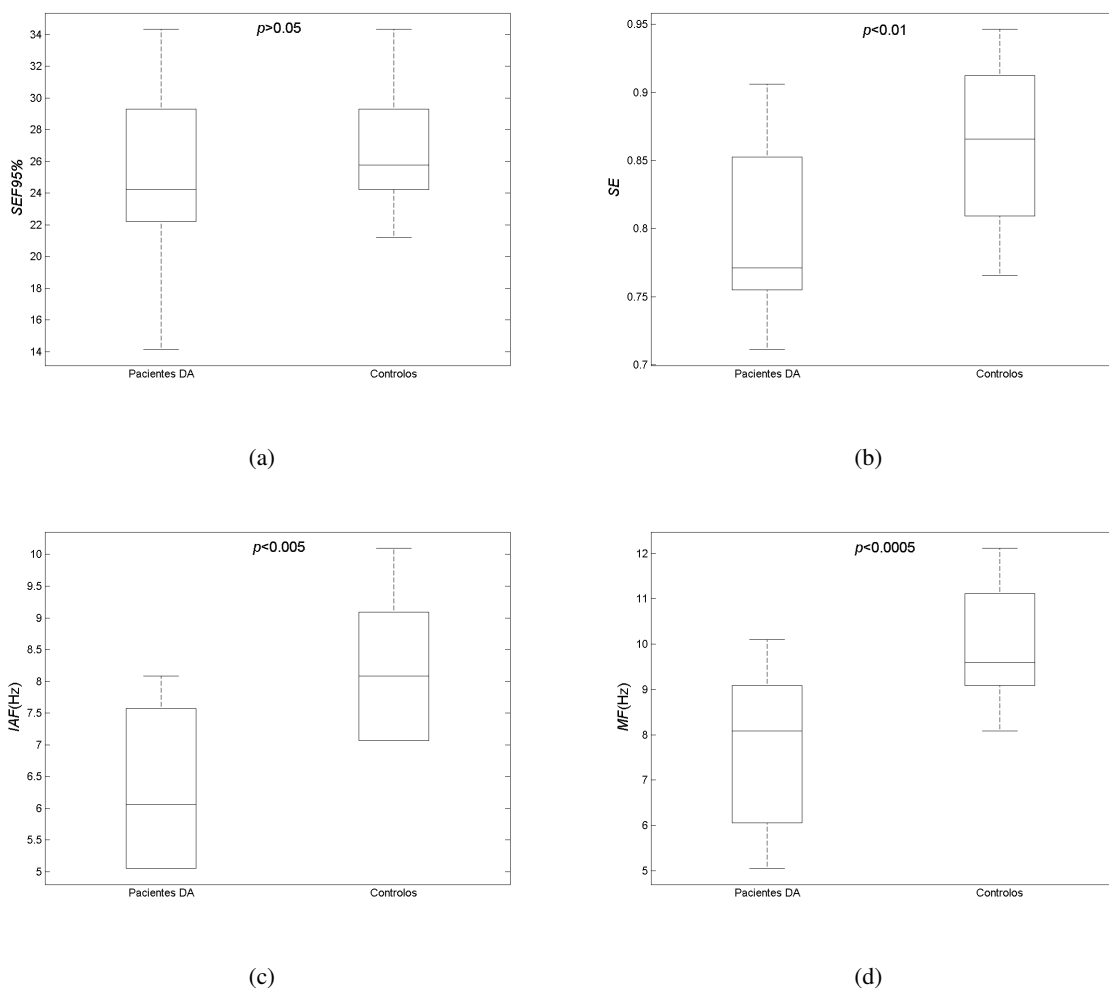
4.3.3.1 Análise estatística / selecção de características

Procedeu-se à análise descritiva das distribuições de parâmetros por indivíduo (pacientes com DA e sujeitos de controlo). Analisou-se a normalidade da distribuição dos parâmetros com o teste de KS e a homocedasticidade pelo teste de Levene.

Tabela 4.28 Teste de KS e teste de Levene das distribuições das *RPs* e dos rcios espectrais.

Parmetros	Teste KS		Teste Levene	
	Rejeita a Hiptese nula	<i>p</i> -valor	Rejeita a Hiptese nula	<i>p</i> -valor
<i>SEF95%</i>	0	0.1753	0	0.3683
<i>SE</i>	1	0.012	0	0,5527
<i>IAF</i>	1	0.0078	0	0,3267
<i>MF</i>	1	0.0067	0	0,0858

A anlise da Tabela 4.28 permitiu concluir que as distribuies de parmetros no atenderam  hiptese dos testes paramtricos. Assim, as diferenas entre os grupos foram estudadas por meio do teste de Kruskal-Wallis por se tratar de um teste no-paramtrico. Foram novamente usados os Diagramas de caixa para se analisarem as diferenas entre as distribuies de parmetros, obtidos pela *PSD_n* por indivduo, dos pacientes de DA e dos sujeitos de controlo.

Figura 4.27 Diagramas de caixa que mostram a distribuio dos valores de alguns parmetros espectrais e correspondentes *p*-valores para todos os canais obtidos pela STFT

Os Diagramas de caixa da Figura 4.27 mostraram que os sujeitos de controlo apresentaram uma média de valores superior aos pacientes com DA em todos os parâmetros, excepto em *SEF95%*. Os parâmetros que exibiram os *p*-valores mais baixos foram o *IAF* e a *MF* (Confrontar Tabela 4.29).

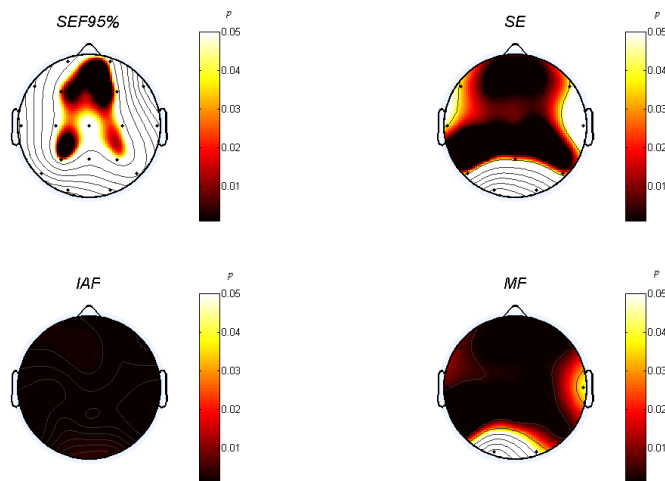
Tabela 4.29 Médias e desvios padrão dos Diagramas de caixa e *p*-valores obtidos pelo Teste de Kruskal-Wallis para a PSD_n da STFT.

Parâmetros	Pacientes com DA Média ±SD	Controlos Média ±SD	<i>p</i> -valores
<i>SEF95%</i>	24.24±5.45	25.75±4.15	0,3003
<i>SE</i>	0.7713±0.06	0.8656±0.056	0,0051
<i>IAF</i>	6.06±1.2	8.08±1.05	0,0017
<i>MF</i>	8.08±1.74	9.596±1.32	0,0003

Tabela 4.30 Resultados de discriminação obtidos através da curva ROC.

Parâmetros	Sensibilidade (%)	Especificidade (%)	Precisão (%)	<i>AROC</i>
<i>SEF95%</i>	60,00%	71,43%	64,71%	0,60
<i>SE</i>	60,00%	92,86%	73,53%	0,79
<i>IAF</i>	55,00%	100,00%	73,53%	0,82
<i>MF</i>	70,00%	85,71%	76,47%	0,85

Figura 4.28 Cabeças que representam as diferenças entre os pacientes com AD e os sujeitos de controlo por eléctrodo com um $p < 0.05$.



A partir da observação da Tabela 4.29 e da Tabela 4.30, verificou-se que os parâmetros *MF* e *IAF* permitiram obter um *p*-valor igual a 0.0003 e 0.0017, respectivamente, e discriminam bem as diferenças entre os dois grupos envolvidos no estudo, dado que os *p*-valores obtidos foram muito baixos. A *SE* também possibilitou uma boa discriminação, ao contrário do *SEF95%*, que apresentou o *p*-valor > 0.05 (o

que equivale a dizer que este parâmetro não apresenta diferenças nas distribuições de valores). Procedeu-se a uma análise mais detalhada das diferenças entre os pacientes com DA e os sujeitos de controlo, neste caso por eléctrodo (Figura 4.28), constatou-se que os parâmetros que apresentaram mais diferenças nos 19 eléctrodos foram os *MF* e *IAF*, os mesmos que apresentavam maiores diferenças por indivíduo. Possibilitando deste modo a realização de uma pré selecção das características que melhor discriminam a DA em relação ao grupo de controlo.

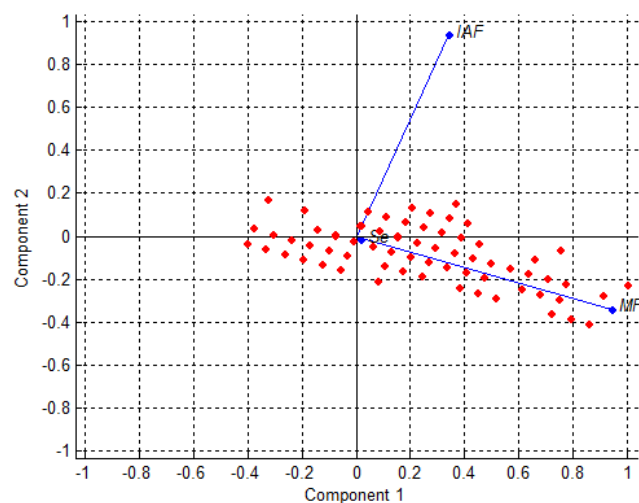
4.3.3.2 Parâmetros espectrais por eléctrodo

Devido à pequena dimensão da BD, nesta abordagem o sinal de cada eléctrodo foi calculado pela média do sinal ao longo dos sucessivos segmentos de 5 segundos de duração - PSD_n por eléctrodo, constituindo um par entrada/saída da RNA e da RL. Os eléctrodos pertencentes ao mesmo indivíduo foram forçados a estar no mesmo conjunto (treino ou de teste).

4.3.3.2.1 Classificação dos sinais EEG

Foram calculadas as duas componentes principais das 3 características (*SE*, *IAF*, *MF*), que apresentaram o p -valor mais baixo quer por indivíduo, quer por eléctrodo, pela técnica de PCA. De forma a poder verificar o modo como cada variável se comportava e contribuiu para as componentes 1 e 2, foram projectados os vectores próprios das características no gráfico xy da Figura 4.29.

Figura 4.29 Biplot das componentes principais das características *SE*, *IAF*, *MF*.



Ao analisar o gráfico da Figura 4.29, verificou-se que a componente principal 2 fez a distinção entre a característica *MF* e a característica *IAF*, pois apresentou um valor positivo e negativo, respectivamente, para as duas características. Os dois vectores representativos de *MF* e *IAF* mostraram um comprimento

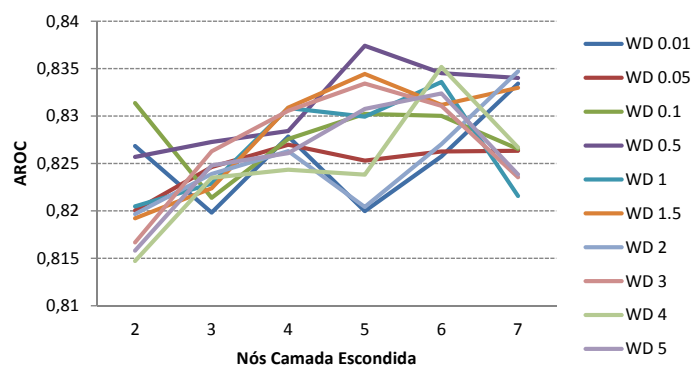
elevado, revelador de que eram as que mais contribuíram para as componentes principais. Apresentaram também uma direcção oposta, e por isso não apresentaram valores sobrepostos, constituindo assim uma boa combinação para a entrada da RNA e da RL. Atendendo a que *SE* exibiu um vector de pequena dimensão contribuiu pouco para as componentes principais podendo induzir em erro a RNA e RL (diminuição do parâmetro *AROC* e da precisão) ou simplesmente não contribuiria significativamente para uma melhoria na discriminação entre a DA e os sujeitos de controlo, quando combinadas com os parâmetros *MF* e *IAF*. Concluindo, a melhor combinação foi dada pelo par *MF* e *IAF* (gráfico da Figura 4.29) e assim, a RNA e a RL apresentaram dois nós de entrada (Tabela 4.31).

Tabela 4.31 Conjuntos de Treino e de Teste.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	
2x494	2x152	2

Dividiram-se os 646 casos no conjunto de treino e de teste. O conjunto de treino ficou constituído por 494 e o de teste por 152 casos (Tabela 4.31). A RNA usada continha a função de activação Logsig, o SCG como algoritmo de treino e a entropia cruzada como a função de erro. Para evitar o treino excessivo da RNA, foi calculado o parâmetro ideal de WD. Para se seleccionarem o melhor parâmetro de WD e o número de nós da camada escondida, usou-se o gráfico da Figura 4.30.

Figura 4.30 Valores de *AROC* ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Feita uma breve análise do gráfico (Figura 4.30), escolheu-se 0.5 de WD e 5 nós na camada escondida, para classificar o conjunto de teste pela RNA, pois estes parâmetros forneceram mais *AROC* pela RNA, para o conjunto de treino, no processo de classificação por validação cruzada deixando um de fora. O conjunto de treino e de teste utilizados para a RNA foram utilizados para a RL. Por fim, os resultados foram avaliados por alguns parâmetros extraídos da curva ROC (Confrontar Tabela 4.32).

Tabela 4.32 Resultados da classificação do grupo de teste

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,86	0,83
Sensibilidade	71,1%	69,8%
Especificidade	97,4%	95,2%
Precisão	84,2%	83,2%

Para verificar o comportamento geral da RL e da RNA com 0.5 de WD e 5 nós na camada escondida, ao conjunto geral de 646 casos procedeu-se ao processo de classificação por validação cruzada deixando um de fora pela RNA e pela RL, e os resultados foram os registados na Tabela 4.33.

Tabela 4.33 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,84	0,81
Sensibilidade	66,3%	64,1%
Especificidade	82,0%	86,8%
Precisão	73,8%	72,9%

4.3.3.3 Parâmetros espectrais - “*grand average*” de 5s

Na abordagem empregue, cada período de 5 segundos do sinal que foi calculado pela média dos 19 eléctrodos por cada 5s, para produzir a chamada “*grand average*” de 5 segundos de duração, correspondendo a um par de entrada/saída da RNA e da RL. As sucessivas “*grand averages*” de 5s por indivíduo foram obrigadas a pertencer ao mesmo conjunto de treino ou de teste.

4.3.3.3.1 Classificação dos sinais EEG

Foram calculadas as componentes principais das 3 características (*SE*, *IAF*, *MF*) que apresentaram o *p*-valor mais baixo quer por indivíduo, quer por eléctrodo, pela técnica de PCA. De forma a poder verificar o modo como cada variável se comportava e contribuiu para as componentes 1 e 2, foram traçados os vectores próprios das características num gráfico xy, da Figura 4.31. Verificou-se que a componente principal 2 fez a distinção entre a característica *MF* e a característica *IAF*, pois apresentou um valor positivo e negativo, respectivamente, para as duas características. Os dois vectores representativos de *MF* e *IAF* apresentaram um comprimento elevado, o que significa que eram as que mais contribuíram para as componentes principais. Mostraram também uma direcção oposta, e por isso não apresentaram valores sobrepostos, constituindo assim uma boa combinação para a entrada da RNA e RL. *SE* exibiu um vector de pequena dimensão pelo que contribuiu pouco para as componentes principais podendo induzir em erro a RNA e a RL (diminuição do parâmetro *AROC* e da precisão) ou simplesmente não contribuiria

significativamente para uma melhoria na discriminação entre a DA e os sujeitos de controlo, quando combinadas com os parâmetros *MF* e *IAF*. Concluindo, a melhor combinação foi dada pelo par *MF* e *IAF* (gráfico da Figura 4.31). Dessa forma a RNA e a RL apresentaram dois nós de entrada (Tabela 4.34).

Figura 4.31 Biplot das componentes principais das características *SE*, *IAF*, *MF*.

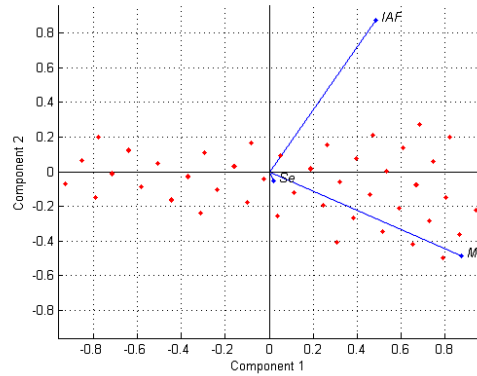
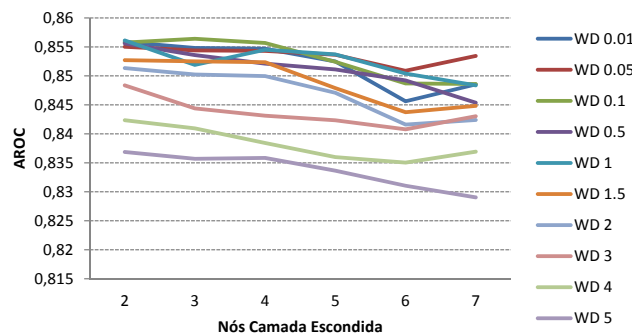


Tabela 4.34 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	
2x780	2x287	2

Dividiram-se os 1067 casos em conjunto de treino e de teste. O conjunto de treino ficou constituído por 780 casos e o de teste por 287 casos (Tabela 4.34). A RNA usada continha a função de activação Logsig, o SCG como algoritmo de treino e a entropia cruzada como a função de erro. Para evitar o treino excessivo da RNA, foi calculado o parâmetro ideal de WD. Para se seleccionarem os melhores parâmetros de WD e o número de nós da camada escondida, usou-se o gráfico da Figura 4.32.

Figura 4.32 Valores de *AROC* ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Foi seleccionado 0.1 de WD e 3 nós na camada escondida. Os resultados foram avaliados por alguns parâmetros extraídos da curva ROC, como se pode ver na Tabela 4.35.

Tabela 4.35 Resultados da classificação do grupo de teste.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,87	0,79
Sensibilidade	92,0%	82,2%
Especificidade	73,6%	61,2%
Precisão	84,0%	69,23%

Para verificar o comportamento geral da RL e da RNA com 0.1 de WD e 3 nós na camada escondida, ao conjunto geral de 1067 casos procedeu-se ao processo de classificação por validação cruzada deixando um de fora pela RNA e pela RL, e os resultados podem ser visto na Tabela 4.36.

Tabela 4.36 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
<i>AROC</i>	0,83	0,81
Sensibilidade	78,9%	79,13%
Especificidade	65,1%	65,1%
Precisão	71,0%	71,1%

4.4 Classificação dos sinais EEG com recurso à DWT

A fim de extrair algumas informações distintas dos sinais de EEG entre os pacientes com DA e os sujeitos de controlo, decompôs-se os segmentos do sinal de 5 segundos pela WT até à decomposição 5 (Confrontar a Figura 4.33).

Foi calculada a Percentagem de Energia (*ER*), correspondente aos coeficientes de detalhe dos níveis 2, 3, 4, 5 e ao coeficiente de aproximação do nível 5. Estes níveis de decomposição correspondem às bandas gama (25-50Hz), beta (12,5-25Hz) alfa (6,25-12,5 Hz), teta (3,125-6,25 Hz) e delta (0-3,125 Hz), respectivamente. Foram calculados os rácios espectrais, conforme mostram as equações seguintes, de forma a explicar a desaceleração do espectro do sinal EEG pela DWT.

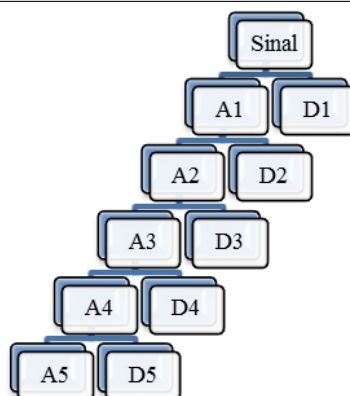
$$e_1 = \frac{ER(alfa)}{ER(teta)}; \quad (4.1)$$

$$e_2 = \frac{ER(alfa) + ER(beta) + ER(gama)}{(ER(teta) + ER(delta))}; \quad (4.2)$$

$$e_3 = \frac{ER(beta)}{ER(delta)}; \quad (4.3)$$

Foram introduzidas pequenas alterações devido às diferenças deste processamento, relativamente ao processamento espectral dado pela *PSD*, impossibilitando a divisão do intervalo de frequências da banda beta em dois (beta1, beta2).

Figura 4.33 Decomposição do sinal pela WT até a decomposição 5



4.4.1 Análise estatística / selecção de características

Foi realizada a análise descritiva da distribuição de parâmetros de pacientes com DA e sujeitos de controlo. Analisou-se a normalidade da distribuição dos parâmetros com o teste de KS e a homocedasticidade pelo teste de Levene. Estes parâmetros foram obtidos pela DWT Bior3.5 por indivíduo. Os resultados são apresentadas na Tabela 4.37.

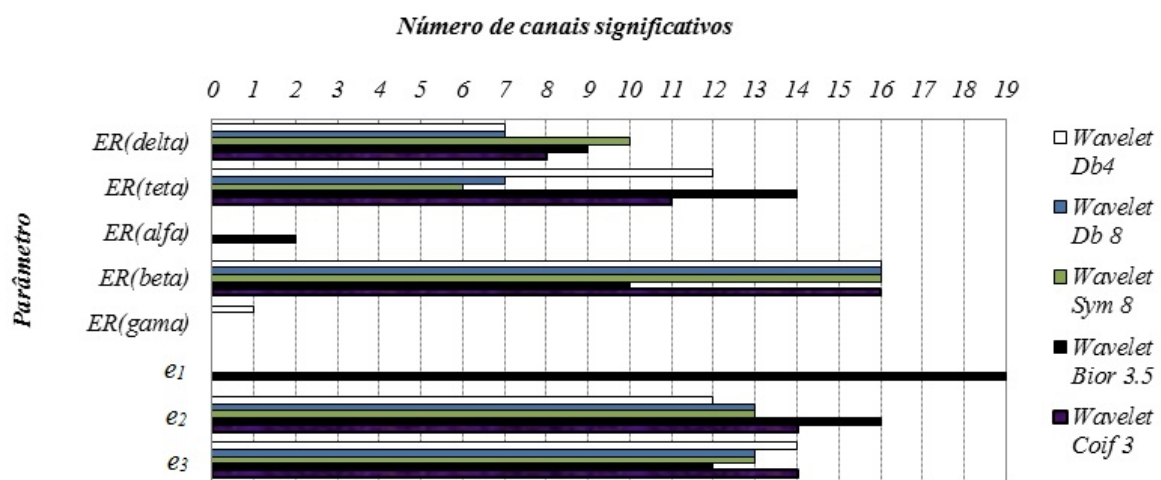
Tabela 4.37 Teste de KS e teste de Levene das distribuições das *RPs* e dos rácios espectrais - por indivíduo.

Parâmetros	Teste KS		Teste Levene	
	Teste Hipótese nula	<i>p</i> -valor	Teste Hipótese nula	<i>p</i> -valor
<i>ER(delta)</i>	1	0,00674	0	0,00069
<i>ER(teta)</i>	1	0,00674	1	0,70995
<i>ER(alfa)</i>	1	0,0078	0	0,04589
<i>ER(beta)</i>	1	0,01191	1	0,30899
<i>ER(gama)</i>	0	0,38122	1	0,57457
<i>e</i> ₁	0	0,00143	1	0,42903
<i>e</i> ₂	1	0,00043	1	0,24305
<i>e</i> ₃	1	0,00086	1	0,91565

No caso do teste de KS, a distribuição de parâmetros obtidos por indivíduo, em geral rejeitou a hipótese nula (normalidade da distribuição dos parâmetros). Isto é, os parâmetros não apresentaram uma

distribuição normal. Por outro lado, no teste de Levene a distribuição de parâmetros, aceitou a hipótese nula em duas distribuições de parâmetros ($ER(\delta)$ e $ER(\alpha)$) o que implica a homocedasticidade das distribuições. Assim, concluiu-se que a distribuição não atendeu à hipótese dos testes paramétricos. Por conseguinte, estudaram-se as diferenças entre os grupos por meio do teste de Kruskal-Wallis que é um teste não-paramétrico. De forma a seleccionar a Wavelet que se ajustou melhor à dinâmica do sinal de EEG nesta metodologia, procedeu-se à comparação entre os resultados dos parâmetros dos pacientes com DA e sujeitos de controlo por eléctrodo, pelo teste de Kruskal-Wallis com a correcção de Bonferroni para $p < 0.05$. Mostra-se de seguida o gráfico que permitiu a selecção da Wavelet a usar.

Figura 4.34 Número de canais significativos para os Processamentos envolvidos neste estudo com p -valores < 0.05 obtidos através do teste de Kruskal-Wallis com correcção de Bonferroni.

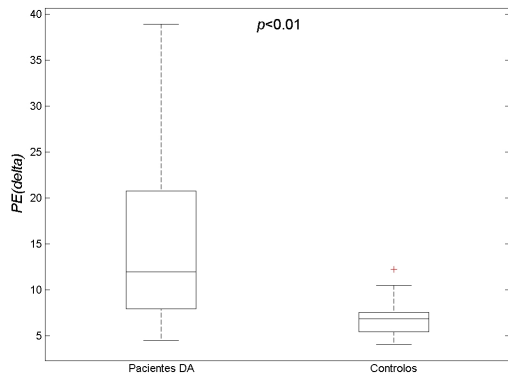


Como pode ser visto na Figura 4.34, a WT Bior 3.5 do sinal foi a que forneceu os canais mais significativos nos rácios espectrais, e_1 e e_2 , relativamente às restantes Wavelets. A Wavelet Bior 3.5 apresentou também parâmetros significativos em alguns canais como no caso da ER em alfa, onde nenhuma outra Wavelet utilizada apresentou. Nas restantes ER s, todas as Wavelets obtiveram os mesmos canais significativos. Finalmente, a Wavelet Db4 apresentou 1 canal com significância em $ER(\gamma)$. Em suma, a Wavelet Bior 3.5 teve mais capacidade em identificar e ressaltar a actividade da doença de Alzheimer do que as restantes Wavelets. A partir deste momento utilizar-se-à a WT Bior 3.5 para classificar os sinais nesta metodologia.

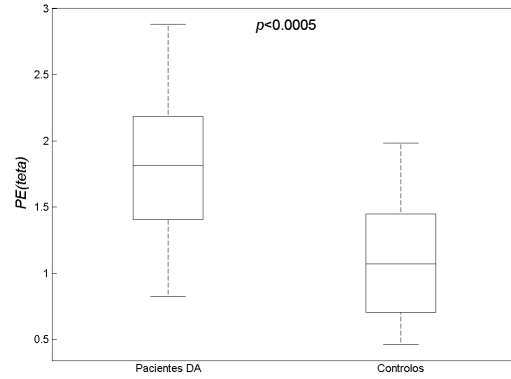
Atendendo os Diagramas de caixa da Figura 4.35 e Figura 4.36, obtidos por parâmetros extraídos por indivíduo, observou-se nos pacientes com DA um aumento da ER nas bandas delta e teta e uma diminuição da ER nas bandas alfa e beta, em comparação com os sujeitos de controlo. Este efeito, chamado de “desaceleração” revelou conformidade com a literatura existente da DA [14, 46]. Pela observação dos valores de ER viu-se que as diferenças mais significativas ($p < 0,001$ e $AROC$ 0,85) foram

obtidos na banda teta. Diferenças significativas também foram obtidas na banda delta ($p < 0,005$), na banda beta ($p < 0,01$) e na banda alfa ($p < 0,01$).

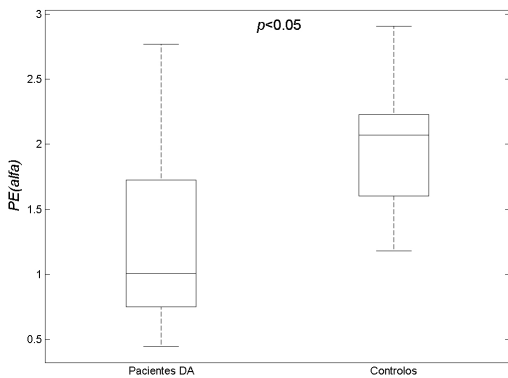
Figura 4.35 Diagramas de caixa que mostram a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais pela WT Bior 3.5.



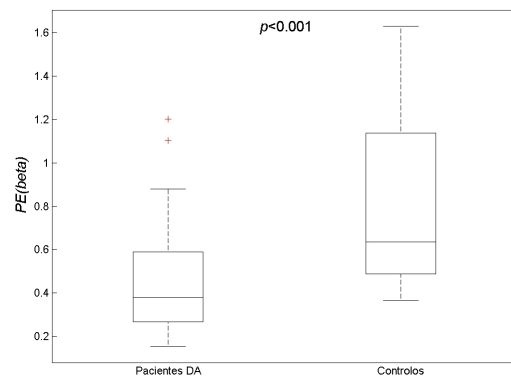
(a)



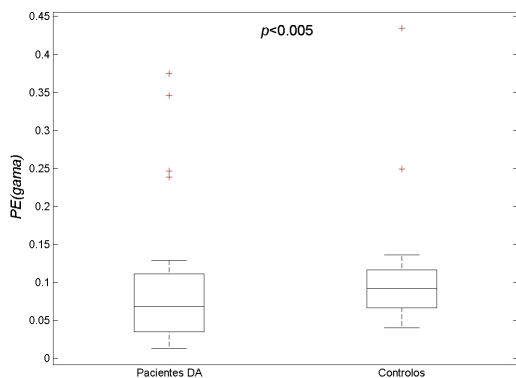
(b)



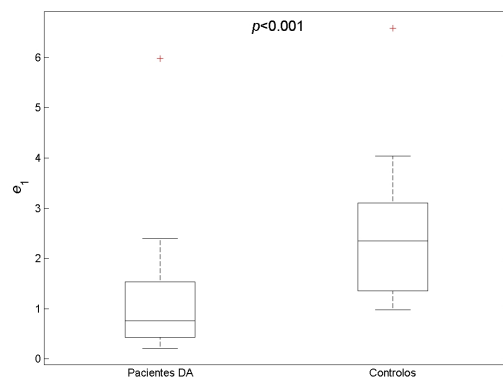
(c)



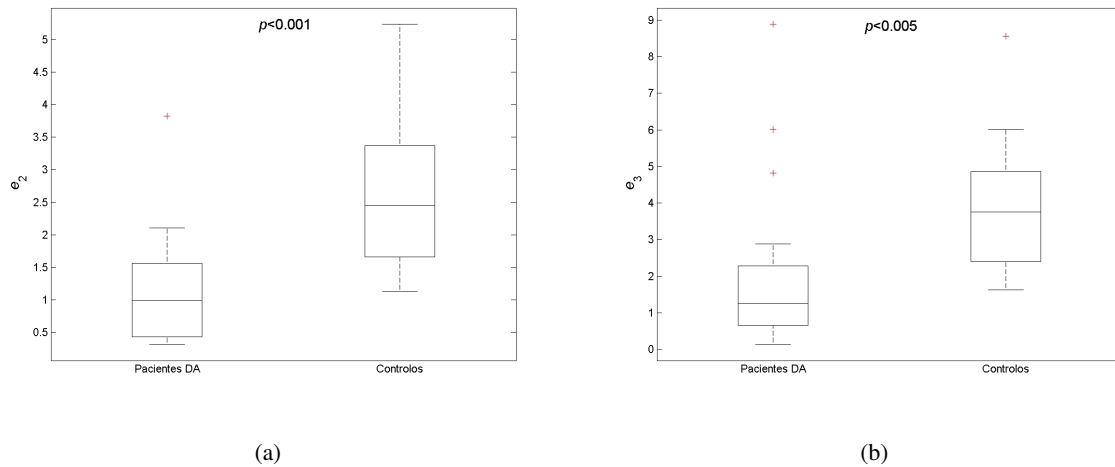
(d)



(e)



(f)

Figura 4.36 Diagramas de caixa que mostram a distribuição dos valores de RP em cada faixa de frequência clássica e respectivos rácios espectrais pela WT Bior 3.5 - Continuação.**Tabela 4.38** Médias e desvios padrão dos Diagramas de caixa e p -valores obtidos pelo Teste de Kruskal-wallis - por eléctrodo.

Parâmetros		Pacientes com DA	Controlos	
		Média \pm SD	Média \pm SD	p -valores
Potências Relativas	$ER(\delta)$	11,91 \pm 9,89	6,84 \pm 2,27	0,00294
	$ER(\theta)$	1,81 \pm 0,52	1,07 \pm 0,46	0,00069
	$ER(\alpha)$	1,01 \pm 0,68	2,07 \pm 0,47	0,00570
	$ER(\beta)$	0,38 \pm 0,30	0,64 \pm 0,37	0,00634
	$ER(\gamma)$	0,07 \pm 0,11	0,09 \pm 0,11	N.S.
Rácios espectrais	e_1	0,76 \pm 1,29	2,35 \pm 1,50	0,00047
	e_2	0,99 \pm 0,87	2,45 \pm 1,11	0,00024
	e_3	1,26 \pm 2,22	3,76 \pm 1,94	0,00207

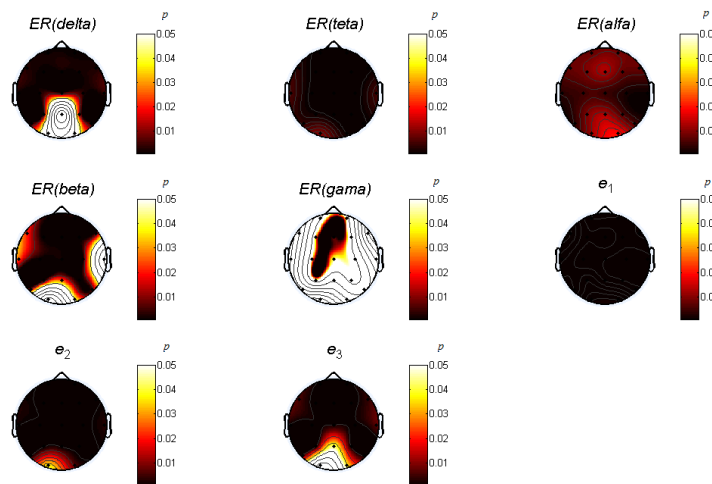
Tabela 4.39 Resultados de discriminação obtidos através da curva ROC, por indivíduo

Parâmetros		Sensibilidade	Especificidade	Precisão	AROC
Potência Relativa	$ER(\delta)$	75,00%	78,57%	76,47%	0,80
	$ER(\theta)$	75,00%	78,57%	76,47%	0,85
	$ER(\alpha)$	75,00%	71,43%	73,53%	0,78
	$ER(\beta)$	65,00%	85,71%	73,53%	0,78
	$ER(\gamma)$	75,00%	50,00%	64,71%	0,63
Rácios espectrais	e_1	75,00%	85,71%	79,41%	0,86
	e_2	75,00%	92,86%	82,35%	0,88
	e_3	65,00%	100,00%	79,41%	0,81

Os resultados ilustrados na Tabela 4.38 e Tabela 4.39 mostraram que os rácios espectrais deram bons resultados ($p < 0,005$ com e_3 e $p < 0,0005$ com o e_1 e e_2 os últimos, com um parâmetro AROC óptimo de 0.86 e 0.88, respectivamente). Fazendo uma análise mais pormenorizada das diferenças de valores

entre os pacientes de DA e dos sujeitos de controlo, neste caso por eléctrodo (Figura 4.37), verificou-se que os parâmetros que apresentaram mais diferenças nos 19 eléctrodos foram os e_1 , e_2 , $ER(teta)$, os mesmos que apresentaram maiores diferenças por indivíduo. Isto, permitiu fazer uma pré-selecção das características que melhor discriminaram os pacientes com DA em relação aos sujeitos de controlo.

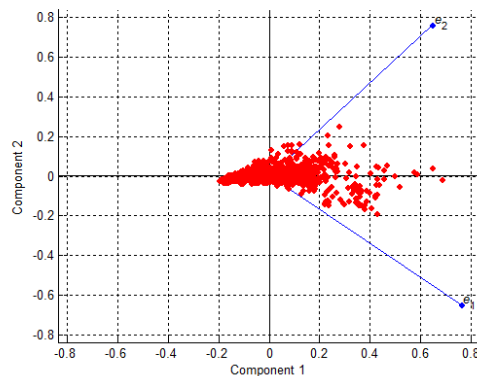
Figura 4.37 Cabeças que representam as diferenças entre os pacientes de Alzheimer e os sujeitos de controlo, por eléctrodo, com um $p < 0.05$



Nesta metodologia assumiu-se que cada eléctrodo constituiria um par entrada/saída da RNA e da RL. Os eléctrodos pertencentes ao mesmo indivíduo foram forçados a estar no mesmo conjunto (treino ou de teste).

Foram calculadas as componentes principais das 3 características (e_1 , e_2 , $ER(teta)$), que apresentaram o p -valor mais baixo por indivíduo e por eléctrodo, pela técnica de PCA. De forma a poder verificar o modo como cada variável se comportava e contribuiu para as componentes 1 e 2, foram projectados os vectores próprios das características num gráfico xy, que pode ser observado na Figura 4.38.

Figura 4.38 Biplot das componentes principais das características e_1 , e_2 , $RP(teta)$.



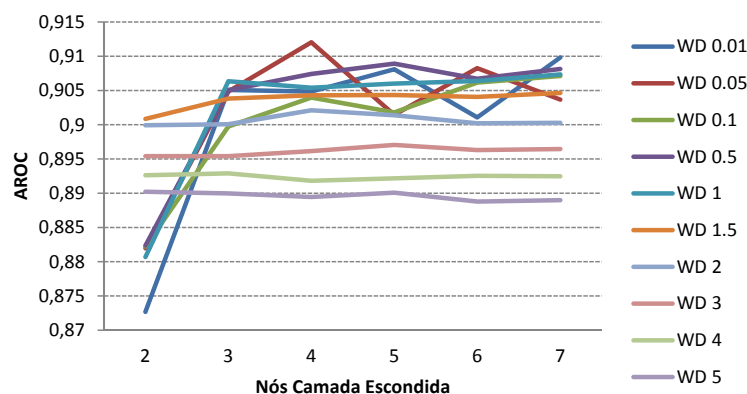
Verificou-se que a componente 2 fez a distinção entre a característica e_1 e a característica e_2 , dado que exibiu um valor positivo e negativo, respectivamente, para as duas características. Os dois vectores representativos de e_1 e e_2 apresentaram um comprimento elevado, revelador de que eram as que mais contribuíram para as componentes principais. Mostraram também uma direcção oposta, o que significa que as duas variáveis não apresentaram valores sobrepostos, constituindo assim uma boa combinação para a entrada da RNA e da RL. A $ER(teta)$ apresentou um vector de dimensão diminuta, pouco contribuiria para as componentes principais e poderia induzir a RNA e a RL em erro (diminuição do parâmetro AROC e da precisão) ou pura e simplesmente não contribuiria significativamente para uma melhoria na discriminação entre a DA e sujeitos de controlo, quando combinadas com os parâmetros e_1 e e_2 . A melhor combinação foi dada pelo par e_1 e e_2 após análise da Figura 4.38 e assim obteve-se uma RNA e um RL com dois nós de entrada (Tabela 4.40).

Tabela 4.40 Conjuntos de Treino e de Teste.

Dimensão		Nós de entrada
Conjunto de treino	Conjunto de teste	
2x494	2x152	2

Dividiram-se os 646 casos no conjunto de treino e de teste. O conjunto de treino ficou constituído por 494 e o de teste por 152 casos (Tabela 4.40). A RNA usada continha a função de activação Logsig, o SCG como algoritmo de treino e a entropia cruzada como a função de erro. Procurando evitar o treino excessivo da RNA, calculou-se o parâmetro ideal de WD. Para se seleccionar o melhor WD e o número de nós da camada escondida, usou-se o gráfico da Figura 4.39.

Figura 4.39 Valores de AROC ao longo de vários nós na camada escondida para diferentes tipos de valores WD no processo de validação cruzada deixando um de fora, do conjunto de treino, da RNA.



Após uma breve análise do gráfico (Figura 4.39), escolheu-se 0.05 de WD e 4 nós na camada escondida. Finalmente, os resultados foram avaliados por alguns parâmetros extraídos da curva ROC.

Tabela 4.41 Resultados da classificação do grupo de teste.

Parâmetros	Método	
	RNA	RL
AROC	0,9	0,8
Sensibilidade	77,6%	77,6%
Especificidade	90,8%	73,7%
Precisão	84,2%	75,7%

Para verificar o comportamento geral da RL e da RNA com 0.05 de WD e 4 nós na camada escondida, ao conjunto geral de 646 casos procedeu-se ao processo de validação cruzada deixando um de fora pela RNA e pela RL, e os resultados foram os seguintes.

Tabela 4.42 Resultados da validação cruzada deixando um de fora sobre o conjunto total de casos.

Parâmetros	Método	
	RNA	RL
AROC	0,87	0,79
Sensibilidade	76,3%	70,0%
Especificidade	82,3%	72,8%
Precisão	78,8%	71,5%

Discussão dos resultados

Com base nos resultados obtidos, discutem-se as diferenças entre os métodos usados e comparam-se com estudos prévios do EEG na DA. No capítulo 2 foi apresentado um novo método baseado na WT para a realização de uma análise espectral. Este método permitiu obter a *PSD* baseada na WT do sinal EEG no domínio da frequência, possibilitando assim um processo de extração de características semelhante ao que acontece com a *PSD* baseada na STFT. Consegue-se desse modo uma comparação com mais precisão entre as técnicas convencionais baseadas na STFT e as técnicas baseadas na WT.

5.1 Aspectos gerais

Pela observação das Tabelas de resultados apresentados pelas três metodologias utilizadas neste estudo, mais concretamente os resultados obtidos pelo Teste de Krukall-Wallis por individuo e por eléctrodo, a classificação dada pelas RNAs e o processo de classificação por validação cruzada da RNA deixando um de fora do conjunto de treino, pode concluir-se que a *PSD* baseada na WT e a DWT apresentaram melhores resultados do que a STFT. Isto talvez tenha ocorrido devido ao facto de a WT permitir mais resolução em frequência do que a STFT. Saliente-se que a WT só depende da escala (frequência) porém, a STFT depende da escala (frequência) e do domínio do tempo. Naturalmente, tem de se chegar a um consenso de modo a que a STFT permita obter informação em ambos os domínios.

Para a criação da *PSD* baseada na WT utilizou-se a WPT, por ser capaz de descrever os locais oscilatórios dos sinais de EEG com mais eficiência do que a STFT e a DWT. É provável que este facto associado à capacidade de particionamento da WPT, que consiste no cálculo simultâneo dos coeficientes de baixa frequência e alta frequência ao longo das várias decomposições do sinal, justifiquem a ligeira diferença de resultados verificada entre a metodologia da *PSD* baseada na WT e a metodologia da DWT.

Neste estudo a *SE* nas duas metodologias associadas à técnica de *PSD* apresentou valores inferiores nos pacientes com DA, relativamente aos sujeitos de controlo, permitindo dizer que os sinais de EEG

dos pacientes com DA foram mais regulares do que os dos sujeitos de controlo. Estes resultados estão em conformidade com os obtidos por Poza [62] e Abásolo [84].

Nos Diagramas de Caixa de cada uma das metodologias observou-se um aumento na RP/ER de pacientes com DA nas bandas delta e teta e uma diminuição da RP/ER nas bandas alfa e beta, em comparação com os sujeitos de controlo. Este efeito, conforme já mencionado noutros capítulos, chamado “desaceleração” é consistente com a literatura existente da DA [14, 46]. Uma hipótese possível para explicar o conceito de “desaceleração do EEG” envolve uma perda do neurotransmissor acetilcolina. Agnoli [13] relatou que após uma administração de drogas colinérgicas a memória dos pacientes com DA apresentou melhorias e os sinais de EEG dos pacientes mostraram uma tendência de mudança para valores normais. Também verificou que o efeito oposto ocorria após uma administração de medicamentos anticolinérgicos. Apesar de estudos indicarem que a perda de acetilcolina está relacionada com o efeito de “desaceleração” do EEG, novos estudos são necessários de modo a confirmar todas as possibilidades.

A DWT e a PSD baseada na WT apresentaram significâncias estatísticas nas bandas delta e alfa, o que não ocorreu com a STFT. Nas bandas teta e beta todos os processamentos apresentaram significância estatística, permitindo assim concluir que a WT e a DWT têm uma forte probabilidade de ter uma capacidade discriminatória superior à STFT.

A DWT e a PSD baseada na WT revelaram maiores significâncias (p -valores) em termos de rácios espectrais relativamente à STFT. Convém salientar que as diferenças entre os pacientes com DA e os sujeitos de controlo se mostraram muito significativas nas três metodologias, especialmente nos rácios espectrais. Por isso, e pelo facto de se observarem diferenças significativas acentuadas em todo o cérebro, em todas as metodologias envolvidas neste estudo, depreende-se que os pacientes se encontravam num estágio avançado da DA. As tendências mantiveram-se em termos de parâmetros de $AROC$ entre as três metodologias. Convém realçar que um p -valor por exemplo de 0.01 não pressupõe uma significância inferior a um p -valor igual a 0.001. O p -valor, apenas serviu de referência na pré selecção das características, apresentadas à entrada da RNA e da RL, que têm uma maior probabilidade de discriminar com eficiência os EEGs entre pacientes com DA e sujeitos de controlo.

As RNAs apresentaram na generalidade, resultados superiores aos da RL, com a excepção registada na metodologia da PSD baseada na STFT, mais concretamente nos Parâmetros espectrais - “*grand average*” de 5s (Confrontar Tabela 4.26), onde a RL na classificação do grupo de Teste conseguiu resultados ligeiramente melhores. Todavia, como os resultados de classificação do grupo de Teste pela RL não tiveram tanta significância como os resultados de classificação dos vários grupos de teste pelas RNAs, na generalidade das metodologias, pode concluir-se que a RL apresentou um desempenho muito inferior relativamente às RNAs. Por conseguinte, será apenas pertinente discutir os resultados de classificação

dada pelas RNAs.

Tabela 5.1 Tabela resumo dos resultados de classificação das Redes Neurais obtidos nas três Metodologias, utilizando as *RP*s/*ER*s e rácios espectrais.

Parâmetros	M. WT				M. STFT				M. DWT	
	P.R.E.E.		P.R.G.A.		P.R.E.E.		P.R.G.A.		T.C.	C.V.
Tipo de classificação	T.C.	C.V.	T.C.	C.V.	T.C.	C.V.	T.C.	C.V.	T.C.	C.V.
<i>AROC</i>	0.97	0.87	0.90	0.85	0.83	0.83	0.81	0.83	0.90	0.87
Sensibilidade	92.1%	78.2%	87.0%	74.8%	76.3%	69.9%	73.6%	74.0%	77.6%	76.3%
Especificidade	90.8%	82.0%	88.0%	77.6%	92.1%	75.5%	90.7%	80.4%	90.8%	82.3%
Precisão	91.5%	79.7%	87.5%	76.1%	84.1%	73.2%	83.3%	76.8%	84.2%	78.8%

M. WT - Metodologia da *PSD* baseada na WT; M. STFT - Metodologia da *PSD* baseada na STFT; M. DWT - Metodologia baseada na DWT; T.C. - Classificação do Grupo de Teste; C.V. - Processo de validação cruzada da RNA deixando um de fora, utilizando todos os dados de EEG; P.R.E.E. - Parâmetros *RP* e rácios espectrais por eléctrodo; P.R.G.A. - Parâmetros *RP* e rácios espectrais - “*grand average*” de 5s;

Tabela 5.2 Tabela resumo dos resultados de classificação das Redes Neurais obtidos nas duas metodologias baseadas na *PSD*, utilizando alguns parâmetros espectrais.

Parâmetros	M. WT				M. STFT			
	P.E.E.E.		P.R.G.A.		P.R.E.E.		P.E.G.A.	
Tipo de classificação	T.C.	C.V.	T.C.	C.V.	T.C.	C.V.	T.C.	C.V.
<i>AROC</i>	0.93	0.87	0.9	0.85	0.86	0.84	0.87	0.83
Sensibilidade	77.6%	75.5%	88.0%	77.6%	71.1%	66.3%	92.0%	78.9%
Especificidade	92.1%	82.0%	87.0%	74.8%	97.4%	82.0%	73.6%	65.1%
Precisão	84.9%	78.2%	87.5%	76.1%	84.2%	73.8%	84.0%	71.0%

M. WT - Metodologia da *PSD* baseada na WT; M. STFT - Metodologia da *PSD* baseada na STFT; T.C. - Classificação do Grupo de Teste; C.V. - Processo de validação cruzada da RNA deixando um de fora, utilizando todos os dados de EEG; P.R.E.E. - Parâmetros espectrais por eléctrodo; P.R.G.A. - Parâmetros espectrais - “*grand average*” de 5s;

As Tabelas 5.1 e 5.2 resumem os resultados de classificação obtidos pelas RNAs das três metodologias envolvidas neste estudo. Pela sua análise verificou-se a existência de diferenças nas três metodologias. Constatou-se uma diferença significativa nos resultados obtidos pela metodologia da *PSD* baseada na WT - Utilização dos parâmetros *RP* e rácios espectrais por eléctrodo, pois conseguiu notoriamente melhores resultados do que as outras duas metodologias. Porém, registou-se uma excepção no que respeita à especificidade, onde a Metodologia da *PSD* baseada na STFT - Utilização dos parâmetros espectrais por eléctrodo teve 97,4%. Os melhores resultados na classificação do conjunto de teste foram conseguidos na metodologia da *PSD* baseada na WT - Utilização dos parâmetros *RP* e rácios espectrais por eléctrodo com 0.97 de *AROC*, 92.1% de sensibilidade e 91.5% de precisão. Os melhores resultados de classificação pelo processo de validação cruzada da RNA deixando um de fora utilizando os dois conjuntos de teste e treino, também foram conseguidos pela metodologia da *PSD* baseada na WT - Utilização dos parâmetros *RP* e rácios espectrais por eléctrodo com 0.87 de *AROC*, 78.2% de sensibilidade, 82.0% de especificidade e 79.7% de precisão.

Como se referiu noutros capítulos, especialmente no capítulo 4, a BD apresentava uma dimensão pequena. Assim, numa tentativa de procurar contornar essa limitação, assumiu-se que cada média dos 19 eléctrodos por cada 5 segundos (“*grand average*” de 5s) e cada média dos segmentos por cada 5s de duração (PSD_n por eléctrodo), constituíram um par entrada/saída da RNA e da RL. Mas isso, conduziria a uma pequena perda de algumas informações espaciais durante o cálculo da média. Perante os resultados constantes nas Tabelas 5.1 e 5.2 verificou-se que os melhores resultados de classificação, muito provavelmente porque houve uma menor perda de informação, foram conseguidos quando se utilizou a PSD_n por eléctrodo como par entrada/saída da RNA. Outra possível justificação para a diferença de resultados pode estar no facto de alguns segmentos de 5 segundos de duração do sinal, poderem não conter informação suficiente para a detecção da DA.

Geralmente o uso de apenas dois nós na camada de entrada da RNA pode ser escasso para as capacidades de uma RNA normal [71]. Todavia, as RNAs utilizadas neste estudo eram especializadas em reconhecimento de padrões com a entropia cruzada como função de erro, que é otimizada para a discriminação de dois grupos distintos com apenas alguns nós na camada de entrada, revelando utilidade na resolução do problema em estudo.

Os resultados conseguidos neste estudo foram na generalidade, muito satisfatórios. Atendendo a que à partida os pacientes que participaram no estudo eram pacientes com probabilidade de terem DA, e dado que se trata de uma doença com diagnóstico clínico de difícil precisão, não havia a certeza absoluta de que os pacientes envolvidos neste estudo tivessem efectivamente DA. Por isso, considera-se que o método da PSD baseada na WT deste estudo pode vir a ser um método de processamento válido para o diagnóstico da DA. Também pode vir a servir de estímulo para mais estudos prospectivos utilizando a WT. Este estudo apresentou resultados promissores e semelhantes aos obtidos por Viallate [15], mas com métodos de processamento distintos e uma diferente BD.

Em trabalhos futuros, deverão ser recolhidos mais sinais EEG para garantir a generalização.

Conclusão

A doença de Alzheimer é uma das doenças mais invalidantes das sociedades desenvolvidas. A importância deste estudo prende-se com o contributo do EEG como um exame complementar de diagnóstico, de modo a possibilitar um diagnóstico diferencial da DA, cada vez mais correcto e precoce.

Neste trabalho avaliou-se a capacidade dos vários métodos de processamento tempo-frequência para caracterizar as propriedades espectrais da actividade EEG espontânea em pacientes com Alzheimer e sujeitos de controlo. Para isso pesquisaram-se e seleccionaram-se métodos de análise tempo-frequência, bem como métodos clássicos espectrais, tais como a STFT. Seguidamente, adaptaram-se os métodos seleccionados ao estudo da actividade do EEG. Modificaram-se e implementaram-se os algoritmos correspondentes a cada medida para avaliar e comparar o seu desempenho ao caracterizar as propriedades dos sinais de EEG de ambos os grupos. Por fim, deduziram-se os padrões normais e patológicos da actividade espontânea do EEG dos pacientes com diagnóstico de DA e sujeitos de controlo para os métodos onde se obtiveram resultados significativos. Ou seja, delineou-se um esquema de trabalho que começou com uma análise abrangente dos padrões espectrais do EEG na DA mediante a STFT. Os resultados obtidos constituíram um ponto de referência para estender os métodos a outros tipos de processamento de sinal baseados na WT. Finalmente aplicaram-se vários processamentos com base na WT para avaliar a sua capacidade em comparação com os métodos clássicos espectrais.

Analisou-se a actividade espontânea dos EEGs de 20 pacientes com diagnóstico de DA (idades compreendidas entre os 58 e 91 anos) e 14 sujeitos de controlo (idades compreendidas entre 61 e 87 anos) e conclui-se que o EEG permite objectivar a presença de demência através da desaceleração do registo por aumento da actividade lenta, nas bandas Delta e Teta, e diminuição da actividade nas bandas Alfa e Beta. Os resultados também evidenciam que a RNA teve uma excelente capacidade de discriminação entre pacientes com DA e grupo de controlo, e sugerem que os parâmetros utilizados podem ser úteis para caracterizar a dinâmica do cérebro associada a esta demência.

Não pode deixar de se referir que este estudo deu o seu contributo para ajudar no diagnóstico da DA, dado que representa um novo esforço para explorar a capacidade da WT no processamento do sinal EEG.

Os resultados mostram que a Wavelet, WT Biortogonal especialmente a Bior 3.5, foi capaz de descrever a dinâmica do EEG com maior poder discriminativo do que a STFT. Os resultados significativos obtidos sugerem que a WT pode constituir uma importante ferramenta para o diagnóstico da DA.

Depois de apresentadas as conclusões mais relevantes expõem-se de seguida as futuras linhas de investigação que surgiram a partir da pesquisa desenvolvida:

1. Extração de parâmetros espectrais dos sinais EEG por técnicas de Fourier e de Wavelet como máximos, mínimos, entropias (Shannon, Energia Logarítmica, sure, entre outras), frequência de transição, linear de Donoho e outros detectores de variações bruscas dos sinais em amplitude, bem como os utilizados neste trabalho (*RP* em cada uma das bandas de frequência, rácios espectrais (r_1, r_2, r_3, r_4), *IAF*, *SEF95%*, *SE* e *MF*).
2. Aumento da base de dados de EEG para permitir uma generalização dos resultados, através de protocolos com clínicas ou hospitais.
3. Criação de um sistema automático, através das Redes Neurais Artificiais, de discriminação de várias doenças, nomeadamente: Alzheimer, Parkinson, Epilepsia, Demência Vascular, Esquizofrenia, ou outras (dependendo do número de Sinais de EEG disponíveis durante a realização de trabalhos futuros).
4. Classificação por grupos etários de modo a descrever o desenvolvimento de doenças Neurológicas (Alzheimer, Parkinson, Epilepsia) e verificar a precisão do diagnóstico do processo em cada um dos níveis etários.
5. Estudo das zonas cerebrais (occipital, frontal, parietal, temporal) mais relevantes para cada uma das doenças envolvidas e para o grupo de controlo, de modo a identificar a zona de maior importância em cada uma das doenças, seleccionando assim os eléctrodos onde se pode identificar com maior precisão a doença em causa.
6. Aplicação de técnicas não lineares, de forma a poder extrair mais características do sinal de EEG, possibilitando assim uma maior diversidade de parâmetros de modo a poder aumentar a precisão e performance das RNAs.

Alguns resultados obtidos neste estudo foram publicados no Symposium Japan - Portugal Nano - Biomedical Engineering Symposium 2011, Bragança, Portugal, 2011 [85]. Foi também aceite um artigo para publicação nos procedings no Workshop ISI HCIST 2011 que se realiza em Outubro 2011 em Vila Moura, Algarve - Portugal [86]. Outros artigos encontram-se em fase de preparação.

Referências

- [1] BLENNOW'S, K. *Amesterdão: European college of neuropsychopharmacology*. ECNP, 2005.
- [2] KNOPMAN, D.; DEKOSKY, S.; CUMMINGS, J.; CHUI, H.; COREY-BLOOM, J. R. Practice parameter: diagnosis of dementia (an evidence-based review). report of the quality standards subcommittee of the american academy of neurology. *Neurology*, v. 56, p. 1143–1153, 2001.
- [3] GABOR, D. Theory of communication. *Journal of the Institute of Electrical Engineers*, v. 93, p. 429–457, 1946.
- [4] COHEN, L. Generalized phase-space distribution functions. *Journal of Mathematical Physics*, New York, v. 7, n.º 5, p. 781–786., 1966.
- [5] BOUDREAUX-BARTELS, G.; MURRAY, R. *The biomedical engineering handbook, 3ª edn*. CRC Press, 2006. Cap. Medical Devices and Systems. Time-Frequency Signal Representations for Biomedical Signals, p. 4–22.
- [6] LI, C.; ZHENG, C. Qrs detection by wavelet transform. *Proceedings of the 15th Annual International Conference of IEEE Engineering in Medicine and Biology Society (IEEE EMBC 1993)*, p. 330–331., 1993.
- [7] BULGRIN, J. R.; RUBAL, B. J.; THOMPSON, C. R.; MOODY, J. M. Comparison of short-time fourier transform, wavelet and time-domain analyses of intracardiac sounds. *Biomedical Sciences Instrumentation*, v. 29, p. 465–472, 1993.
- [8] LIN, Z. Y.; CHEN, J. D. Z. Time-frequency representation of the electrogastrogram - application of the exponential distribution. *IEEE Transactions on Biomedical Engineering*, v. 41, p. 267–27, 1994.
- [9] EGGERMONT, J. J.; SMITH, G. M. Characterizing auditory neurons using the wigner and rihaczek distributions: A comparison. *Journal of the Acoustical Society of America*, v. 87,nº1, p. 246–259, 1990.
- [10] SCHIFF, S. J.; MILTON, J. G. Wavelet transforms for electroencephalographic spike and seizure detection. *Proceeding of the SPIE International Society for Optical Engineering*, v. 2036, p. 50–56, 1993.
- [11] PETROSIAN, A.; PROKHOROV, D.; LAJARA-NANSON, W.; SCHIFFER, R. Recurrent neural network-based approach for early recognition of alzheimer's disease in eeg. *Clinical Neurophysiology*, v. 112, n.º 8, p. 1378–1387, 2001.
- [12] POLIKAR, R.; TOPALIS, A.; GREEN, D.; KOUNIOS, J.; CLARK, C. Comparative multiresolution wavelet analysis of erp spectral bands using an ensemble of classifiers approach for early

- diagnosis of alzheimer's disease. *Computers in Biology and Medicine*, v. 37, n.º 4, p. 542–558, 2007.
- [13] AGNOLI, A.; MARTUCCI, N.; MANNA, V.; CONTI, L.; FIORAVANTI, M. Effect of cholinergic and anticholinergic drugs on short-term memory in alzheimer's dementia: a neuropsychological and computerized electroencephalographic study. *Clinical Neuropharmacology*, v. 6, p. 311–23, 1983.
- [14] STAM, C. Nonlinear dynamical analysis of eeg and meg: Review of an emerging field. *Clinical Neurophysiology*, v. 116, p. 2266–2301, 2005.
- [15] VIALATTE, F.; CICHOCKI, A.; DREYFUS, G.; MUSHA, T.; RUTKOWSKI, T.; BLIND, G. R. Early detection of alzheimer's disease by blind source separation and bump modelling of eeg signals. *Lectures Notes in Computer Science - Springer*, v. 3596, p. 683–692, 2005.
- [16] SUBASI, A.; ERÇELEBI, E. Classification of eeg signals using neural network and logistic regression. *Computer Methods and Programs in Biomedicine*, v. 78, p. 87–99, 2005.
- [17] GUO, L.; RIVERO, D.; SEOANE, J.; PAZOS, A. Classification of eeg signals using relative wavelet energy and artificial neural networks. *GEC Summit*, p. 177–184, 2009.
- [18] GAWEL, M.; SZMIDT-SALKOWSKA, E.; KOWALSKI, J. The value of quantitative eeg in differential diagnosis of alzheimer's disease subcortical vascular dementia. *J Neurol Sci*, v. 283, p. 127–133, 2009.
- [19] RODRIGUES, P.; TEIXEIRA, J. Classification of electroencephalogram signals using artificial neural networks. *3rd International Conference on Biomedical Engineering and Informatics*, v. 2, n. ISBN-978-1-4244-6496-8, p. 808–812, 2010.
- [20] NIEDERMEYER, E.; LOPES DA SILVA, F. *Fisiologia*. Elsevier, 1999. Cap. Electroencephalography: Basic principles, clinical applications, and related fields, p. 2391–2399.
- [21] SAEID SANEI, J. A. *Eeg signal processing*. Wiley, 2007.
- [22] BERNE, R.; LEVY, M. *Fisiologia*. Elsevier, 2004.
- [23] SCHÜRMAN, M.; BASAR, E. Functional aspects of alpha oscillations in the eeg. *Int J Psychophysiol*, v. 39, p. 2–3, 2001.
- [24] ZSOLT, L. K. *Redes neurais artificiais - fundamentos e aplicações*. Livraria Da Fisica, 2002.
- [25] FRANÇA, A. I. Event related brain potentials. [Acedido a 24 de Novembro de 2010]. Disponível na Internet: www.acesin.letras.ufrj.br/curso_neurofisiologia/eeg.ppt.
- [26] KANDEL, E.; SCHWARTZ, J. H.; JESSELL, T. M. *Principles of neural science*. McGraw-Hill, 2000.
- [27] BRODMANN, K.; GAREY, L. J. *Brodmann's: Localisation in the cerebral cortex*. Springer, 2005.
- [28] PARRA, J.; SILVA, L. D. *Manual da electroencefalografía*. McGraw-Hill Interamericana, 2001. Cap. Bases neurofisiológicas de la electroencefalografía y magnetoencefalografía, p. 1–17.
- [29] CASTRO, C.; ÁLVARO, P. *A doença de alzheimer e outras demências em portugal*. Lidel, 2005.
- [30] RIEMENSCHNEIDER, M.; LAUTENSCHLAGER, N.; WAGENPFEIL, S., D. J.; DRZEZGA, A.; KURZ, A. Cerebrospinal fluid tau and beta-amyloid 42 proteins identify alzheimer disease in subjects with mild cognitive impairment. *Arch Neurol.*, v. 59, p. 29–34, 2002.

- [31] ASSOCIATION, A. Viagem ao cérebro. [Acedido a 24 de Novembro de 2010]. Disponível na Internet: http://www.alz.org/brain_portuguese.
- [32] SEELEY; STEPHENS; TATE. *Anatomia & fisiologia 6ª edição*. Lusodidacta, 2005.
- [33] BLUMENAU, U. Doença de alzheimer. [Acedido a 3 de Janeiro de 2011]. Disponível na Internet: https://www.furb.br/especiais/download/760525-933424/Doenca_Alzheimer.pdf.
- [34] BALLONE, G. Psiqweb - demências. [Acedido a 21 de Março de 2011]. Disponível na Internet: www.lactobacilo.com/dietoterapia/alzheimer.htm.
- [35] PANISSET, M.; STERN, Y. . Clinical diagnosis and management of alzheimer's disease. *Martin Dunitz Publishers*, p. 129–139, 1996.
- [36] SHANKLE, W.; AMEN, D. *Preventing alzheimer's: prevent, detect, diagnose, treat, and even halt alzheimer's disease and other causes of memory loss*. G. P. Putnam's Sons, 2004.
- [37] JONKER, C.; LAUNER, L.; HOOIJER, C.; LINDEBOOM, J. Memory complaints and memory impairment in older individuals. *J Am Geriatr Soc*, p. 144–149, 1996.
- [38] FELDMAN, H.; WOODWARD, M. The staging and assessment of moderate to severe alzheimer disease. *Neurology*, v. 65, p. 10–17, 2005.
- [39] ORTIZ, K. *Distúrbios neurológicos adquiridos*. Manole, 2005.
- [40] GAY, B.; TAYLOR, K.; HOHL, U.; TOLNAY, M.; STAEHELIN, H. The validity of clinical diagnoses of dementia in a group of consecutively autopsied memory clinic patients. *J Nutr Health Aging*, p. 132–137, 2008.
- [41] JELIC, V.; JOHANSSON, S., A. O.; SHIGETA, M.; JULIN, P.; NORDBERG, A. Quantitative electroencephalography in mild cognitive impairment: longitudinal changes and possible prediction of alzheimer's disease. *Neurobiol Aging*, v. 21, p. 533–540, 2000.
- [42] BERTOLUCCI, P.; OKAMOTO, I.; BRUCKI, S.; SIVIERO, M.; TONIOLO, N. J.; RAMOS, L. Applicability of the cerad neuropsychological battery to brazilian elderly. *Arq Neuropsiquiatr*, v. 59, p. 532–536, 2001.
- [43] CERDEIRA, V. Diário de medicina preventiva, Maio 2007. [Acedido a 4 de Abril de 2011]. Disponível na internet: http://coursejournal_medicina.blogs.sapo.pt/2007/05/.
- [44] MANIDI, M. J.; MICHEL, J. P. *Atividade física para adultos com mais de cinquenta anos*. Manole, 2001.
- [45] RAMOS, L. R.; MACEDO, M. B. Como diagnosticar distúrbios de memória e demência. *Ciber Saúde*, 2001.
- [46] JEONG, J. Eeg dynamics in patients with alzheimer's disease. *Clin. Neurophysiol*, v. 115, p. 1490–1505, 2004.
- [47] SIGNORINO, M.; PUCCI, E.; BELARDINELLI, N.; NOLFE, G.; ANGELERI, F. Eeg spectral analysis in vascular and alzheimer dementia. *Electroencephalogram Clin Neurophysiol*, v. 25, p. 94–313, 1995.
- [48] SOININEN, H.; PARTANEN, J.; LAULUMAA, V.; HELKALA, E. L.; LAAKSO, M.; RIEKKINEN, P. Longitudinal eeg spectral analysis in early stage of alzheimer's disease. *Clin Neurophysiol*, p. 72–290, 1989.

- [49] NISE, N. S. *Control systems engineering*. John Wiley & Sons, 2000.
- [50] COHEN, A. *The biomedical engineering handbook, 2^a edn*. New York: IEEE Press., 2000. Cap. Biomedical Signals: Origin and dynamic characteristics: Frequency- domain analisys, p. 1–52.
- [51] CHALLIS, K. *Biomedical signal processing (time-domain methods)*. Med. & Biol. & Comput, 1990.
- [52] BLANCO, S.; FIGLIOLA, A.; QUIROGA, R.; ROSSO, O.; SRRANO, E. Time-frequency analysis of electroencephalogram series. iii. wavelet packets and information cost function. *Phys. Rev.*, p. 932–940, 1998.
- [53] CARR, J. J.; BROWN, J. M. *Introduction to biomedical equipment technology*. Prentice Hall, 2001.
- [54] CARNEIRO, M. B. P. Disciplina: Processamento de sinais biomédicos, 2009. [Acedido a 24 de Novembro de 2010]. Disponível na Internet: http://www.feelt.ufu.br/pastas/Processamento_de_Sinais_Biomedicos/NotasAulaPDS_Cap1_2_3.pdf.
- [55] HAYKIN, S.; VEEN, B. V. *Signals and systems*. John Wiley & Sons, 2003.
- [56] HAYKIN, S. *Communication systems*. 2001.
- [57] RIOUL, O.; VETTERLI, M. Wavelets and signal processing. *IEEE Signal Processing Magazine*, v. 8, p. 14–38, 1992.
- [58] KAISER, G. *A friendly guide to wavelets*. Birkhäuser Boston, 1994.
- [59] MATHWORKS. About wavelet packet analysis. [Acedido a 23 de Novembro de 2010]. Disponível na Internet: <http://www.mathworks.com/help/toolbox/wavelet/ug/f7-1011118.html>.
- [60] WICKERHAUSER, M. *Acoustic signal compression with wave packets*. Preprint Yale University, 1989.
- [61] POZA, J.; HORNERO, R. *Recent advances in biomedical signal processing*. Bentham Science Publishers, 2010. Cap. Time-Frequency Analysis of MEG activity in Alzheimer’s disease, p. 122–139.
- [62] POZA, J.; HORNERO, R.; ABÁSULO, D.; FERNÁNDEZ, A.; MARIA, G. Extraction of spectral based measures from meg background oscillations in alzheimer’s disease. *Medical Engineering & Physics*, v. 29, p. 1073–1083, 2007.
- [63] POZA, J.; HORNERO, R.; ABÁSULO, D.; FERNÁNDEZ, A.; MAYO, A. Evaluation of spectral ratio measures from spontaneous meg recordings, in patients with alzheimer’s disease. *Computer Methods And Programs In Biomedicine*, v. 90, p. 34–47, 2008.
- [64] HAYKIN, S. *Neural networks*. Prentice Hall, 1998.
- [65] MOTA, G. L. A. Análise da mesoestrutura de bambu utilizando medidas de saliência de textura. [Acedido a 15 de Novembro de 2010]. Disponível na Internet: <http://www.visgraf.impa.br/Courses/ip00/proj/texta/>.
- [66] JONES, E. R. Neural networks’ role in predictive analytics, Fevereiro 2008. [Acedido a 29 de Dezembro de 2010]. Disponível na Internet: http://www.information-management.com/specialreports/2008_61/10000704-1.html.
- [67] PEDREIRA, C. E. *Agrupamento e visualização de dados sísmicos através de quantização vetorial*. 2004. Tese (Doutorado em Física) - Universidade Católica do Rio de Janeiro, 2004.

- [68] LVCON. Redes neurais artificiais. [Acedido a 29 de Dezembro de 2010]. Disponível na Internet: <http://www.natcomp.com.br/lvcon/tema?tema=2>.
- [69] TEIXEIRA, J. P. *Sebenta redes neuronais artificiais : Introdução às redes artificiais neuronais, redes feed-forward em matlab*. Instituto Politécnico de Bragança, 2010.
- [70] MOLLER, M. F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, v. 6, p. 525–533, 1993.
- [71] BISHOP, C. M. *Neural networks for pattern recognition*. Oxford University Press., 1995.
- [72] PRIDDY, K. L.; KELLER, P. E. *Artificial neural networks: An introduction*. SPIE Press Book, 2005.
- [73] VENTICINQUE, E. M.; CARNEIRO, J. S.; MOREIRA, M. P.; FERREIRA, L. O uso de regressão logística para espacialização de probabilidades. *Megadiversidade*, v. 3, p. 25–37, 2007.
- [74] AGRESTI, A. *An introduction to categorical data analysis, 2nd edition*. Wiley, 2007.
- [75] WILLIAMS, C.; LEE, S.; FISHER, R.; DICKERMAN, L. A comparison of statistical methods for prenatal screening for down syndrome. *Applied Stochastic Models in Business and Ind.*, v. 15, p. 89–101, 1999.
- [76] SILVA, A. *Introdução à análise de dados*. E-paper Serviços Editoriais, 2009.
- [77] BAJPAI, N. *Business statistics*. Number ISBN:9788131726020. Pearson Education, 2010.
- [78] HOTELLING, H. *Analysis of a complex of statistical variables into principal components*. Journal of Educational Psychology, 1933. v. 24.
- [79] PEARCE, T. C.; SCHIFFMAN, S. S.; NAGLE, H. T. *Handbook of machine olfaction*. Wiley-VCH, 2003.
- [80] PUCCI, E.; CACCHIÒ, G.; ANGELONI, R.; BELARDINELLI, N.; NOLFE, G.; SIGNORINO, M.; ANGELERI, F. Eeg spectral analysis in alzheimer's disease and different degenerative dementias. *Arch. Gerontol. Geriatr.*, v. 26, p. 283–297, 1998.
- [81] JELIC, V.; BLOMBERG, M.; DIERKS, T.; BASUN, H.; SHIGETA, M.; JULIN, P.; JENSEN, M.; LANNFELT, L. WINBLAD, B.; WAHLUND, L. Eeg slowing and cerebrospinal fluid tau levels in patients with cognitive decline. *Neuroreport*, v. 9, p. 157–160, 1988.
- [82] BENNYNS, K.; RONDOUIN, G.; C., V.; TOUCHON, J. Diagnostic value of quantitative eeg in alzheimer's disease. *Neurophysiol. Clin.*, v. 31, p. 153–160, 2001.
- [83] STEFAN, H.; RAMPP, S. Magnetoencephalography in presurgical epilepsy diagnosis. *Expert Rev Med Devices*, v. 47, p. 335–347, 2007.
- [84] ABÁSULO, D.; HORNERO, R.; P., E.; ÁLVAREZ, D.; POZA, J. Entropy analysis of the eeg background activity in alzheimer's disease patients. *Physiol Meas* 2006, v. 27, p. 41–53, 2006.
- [85] RODRIGUES, P.; TEIXEIRA, J.; HORNERO, R.; POZA, J.; CARRERES, A. Classification of alzheimer's electroencephalograms using artificial neural networks and logistic regression. *Japan - Portugal Nano-Biomedical Engineering Symposium 2011*, v. 1, n. ISBN-4-904157-20-6, p. 33–34, 2011.
- [86] RODRIGUES, P.; TEIXEIRA, J. Artificial neural networks in the discrimination of alzheimer's disease. *HCIST Centeris 2011 -Springer Verbag , Artigo Aceite para publicação Outubro*, 2011.