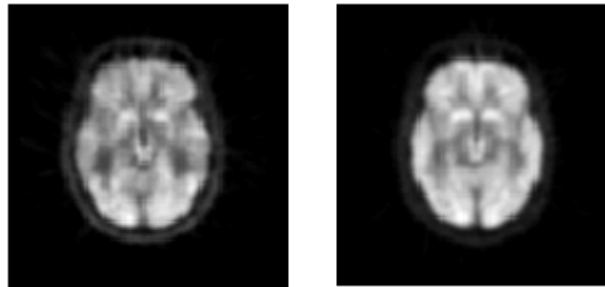




INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa



RECONHECIMENTO DE PADRÕES DA DOENÇA DE ALZHEIMER EM IMAGENS PET

Fábio Joel Vieira da Silva

Dissertação para a obtenção do Grau de Mestre em
Engenharia Electrotécnica e de Computadores

Presidente: Prof. Carlos Jorge Ferreira Silvestre

Orientadora: Prof. Maria Margarida Campos da Silveira

Co-Orientador: Prof. Jorge dos Santos Salvador Marques

Vogal: Prof. Ana Luísa Nobre Fred

Dezembro de 2009

Agradecimentos

Gostaria de agradecer à minha orientadora, Professora Margarida Silveira, pela total disponibilidade durante a realização da minha dissertação de mestrado. Agradeço também o modo como orientou esta dissertação, e pelos seus esclarecimentos que contribuíram para a minha formação.

Quero agradecer ao meu co-orientador, Professor Jorge Salvador Marques, pelo seu apoio, disponibilidade e pelas suas explicações ao longo da elaboração da minha dissertação de mestrado.

Quero agradecer ao Dr. António Bastos Leite, pela sua ajuda na análise e interpretação das imagens PET.

Quero agradecer aos meus camaradas da Academia Militar por todo o apoio que me prestaram ao longo do curso. Sem a vossa ajuda era impossível atingir o fim da primeira etapa. Agradeço a Academia Militar pela sua formação academia, militar e por me ter dotado de princípios e valores, pelos quais eu me oriento.

Quero agradecer aos meus amigos do Porto pelos momentos de laser e descontração. Agradeço a vossa amizade e paciência nos momentos mais críticos.

Quero agradecer à minha Mãe, ao meu Pai e ao meu Irmão por toda atenção e carinho que me transmitiram ao longo da minha vida. Agradeço bastante todo o vosso empenho para me fornecer o bem-estar e as condições para eu ser feliz.

Obrigado a todos.

Resumo

A doença de Alzheimer provoca alterações degenerativas no tecido nervoso, que resultam na progressiva perda de memória e das funções cognitivas. A doença de Alzheimer é o tipo de demência mais frequente. A prevalência da doença de Alzheimer aumenta com idade e, dado o envelhecimento da população, é expectável que a sua incidência aumente. Actualmente, não há cura para a doença de Alzheimer, mas com diagnóstico e tratamento precoces é possível atrasar a sua progressão.

O exame PET indica as alterações funcionais no cérebro humano e, actualmente é uma ferramenta útil na análise do fluxo sanguíneo cerebral. A doença de Alzheimer provoca a redução do fluxo sanguíneo nas regiões propícias à sua propagação, principalmente no lobo temporal, lobo parietal e o hipocampo. Contudo a difícil análise visual e interpretação das imagens PET no diagnóstico da doença de Alzheimer motivou a necessidade da utilização de métodos de reconhecimento de padrões sobre as imagens PET como ferramentas de diagnóstico automático da doença de Alzheimer.

Nesta dissertação foram utilizados os métodos dos k -vizinhos mais próximos e de Boosting na discriminação entre classes AD e Normal, MCI e Normal, e finalmente entre as classes AD e MCI, com o intuito de torná-los numa ferramenta útil e eficaz no diagnóstico automático da doença de Alzheimer. A implementação da classe MCI, raramente estudada, permite diagnosticar a doença de Alzheimer num estado precoce através da distinção binária entre as classes MCI e Normal. Ambos os métodos utilizados na identificação da doença de Alzheimer estão sujeitos ao problema da “maldição da dimensionalidade”.

O método de Boosting, através da variante Adaboost, atingiu melhores desempenhos do que o método dos k -vizinhos mais próximos em todas as discriminações entre as classes. Assim, o método de Boosting adaptou-se melhor ao problema da “maldição da dimensionalidade” e possui sérias potencialidades de se tornar numa ferramenta de diagnóstico automático da doença de Alzheimer.

Palavras-chave: Doença de Alzheimer, PET, diagnóstico automático, k -vizinhos mais próximos, Boosting, Adaboost.

Abstract

Alzheimer's disease causes degenerative changes in nervous the tissue, which results in progressive loss of memory and cognitive functions. Alzheimer's disease is the most frequent type of dementia. The prevalence of Alzheimer's disease increases in the elderly patients, and given the fact that population is getting older it is expected that incidence increases. Currently, there is no cure for Alzheimer's disease, but early diagnosis and treatment can slow its progression.

The PET examination specifies the functional changes in the human brain and it is an useful tool in analyzing the cerebral blood flow. The regional abnormalities for Alzheimer's disease cause the reduction of blood flow, mainly the temporal lobe, parietal lobe and hippocampus. However, the difficulty of the analysis and interpretation of PET imaging in the diagnosis of Alzheimer's disease led to the use of methods of pattern recognition on PET imaging as a tool for automatic diagnosis of Alzheimer's disease.

In this dissertation, we used the k -nearest neighbor and Boosting methods in discrimination between the AD and Normal classes, MCI and Normal, and finally between the AD and MCI classes, with the purpose of making them a useful and effective tool in the automatic diagnosis of Alzheimer's disease. The implementation of the MCI class, rarely studied, allows the early diagnosis of Alzheimer's disease through the binary distinction between the MCI and Normal classes. Both methods are afflicted by the problem of "*curse of dimensionality*".

The Boosting method Adaboost achieved better performance than k -nearest neighbors method in all discrimination between classes. Thus, the Boosting method adapted better to the problem of "*curse of dimensionality*" and has serious potential to become a tool for automatic diagnosis of Alzheimer's disease.

Keywords: Alzheimer's disease, PET, automatic diagnosis, k -nearest neighbor, Boosting, Adaboost.

Índice

Agradecimentos.....	I
Resumo	II
Abstract	III
1. Introdução.....	1
1.1. Doença de Alzheimer	1
1.2. Diagnóstico da Doença de Alzheimer	3
1.3. Estudos Previamente Realizados no Diagnóstico Automático da Doença de Alzheimer.....	3
1.4. Objectivo da Dissertação.....	4
1.5. Contribuições da Dissertação.....	5
1.6. Organização da Dissertação	5
2. Revisão dos Classificadores Utilizados na Identificação da Doença de Alzheimer em Imagens PET e MRI.....	7
2.1. Reconhecimento de Padrões	7
2.1.1. Extração de Características e o Classificador	7
2.1.2. Avaliação do Desempenho Utilizando Validação Cruzada	8
2.2. Reconhecimento de Padrões de Alzheimer nas Imagens PET e MRI.....	8
2.2.1. Imagens PET	9
2.2.2. Imagens MRI.....	10
2.3. Extração de Características das Imagens PET e MRI	11
2.4. Resultados dos Classificadores Utilizados na Detecção da Doença de Alzheimer sobre as Imagens PET e MRI	12
3. Diagnóstico Automático da Doença de Alzheimer	16
3.1. Extração de Características.....	16
3.1.1. Features de Haar	17
3.2. Métodos dos k -Vizinhos Mais Próximos e de Boosting	18
3.2.1. Métodos dos k -Vizinhos Mais Próximos.....	18
3.2.1.1. Exemplo do método dos k -vizinhos mais próximos	20
3.2.2. Método de Boosting	21
3.2.2.1. Algoritmo Hedge.....	22
3.2.2.2. Adaboost	24

3.2.2.2.1. Análise do erro de treino.....	25
3.2.2.2.2. Exemplo do Adaboost.....	26
3.2.2.2.3. Generalização do Adaboost	28
3.2.2.3. Classificação de multi-classes	28
3.2.2.3.1. Adaboost.M1.....	28
3.2.2.3.2. Adaboost.M2.....	29
4. Resultados Experimentais.....	31
4.1. Base de Dados ADNI.....	31
4.2. Pré-processamento das Imagens PET.....	32
4.3. Utilização de Máscaras	33
4.4. Resultados do Método dos k -Vizinhos Mais Próximos	34
4.4.1. Características Fundamentais	34
4.4.2. Classificação	37
4.4.3. Avaliação do Desempenho do Método dos k -Vizinhos Mais Próximos	37
4.4.4. Resultados e Análise	37
4.4.5. Localização das Principais Regiões Indicadas pelo k -Vizinhos Mais Próximos	40
4.5. Resultados do Adaboost	41
4.5.1. Características Fundamentais	41
4.5.2. Classificação	43
4.5.3. Avaliação do Desempenho do Adaboost.....	43
4.5.4. Resultados e Análise	43
4.5.5. Localização das Principais Regiões Indicadas pelo Adaboost.....	46
4.6. Comparação dos Resultados	47
5. Conclusão.....	49
5.1. Trabalho Futuro	50
Anexo A.....	51
Anexo B.....	52
Referências Bibliográficas.....	55

Lista de Figuras

Figura 1.1: Distribuição das demências	1
Figura 1.2: A prevalência da doença de Alzheimer aumenta com a idade	2
Figura 1.3: Cérebro humano: a) Localização dos lobos; b) Localização do hipocampo	2
Figura 2.1: Estrutura de um sistema de reconhecimento de padrões	7
Figura 2.2: Imagem PET em corte axial: a) classe AD; b) classe Normal.....	9
Figura 2.3: Imagem MRI em corte axial: a) classe AD; b) classe Normal	11
Figura 2.4: As localizações das sete áreas utilizadas no treino do classificador	12
Figura 3.1: Features de Haar a) simples; b) extensões	17
Figura 3.2: Classificação do método dos k-vizinhos mais próximos	21
Figura 3.3: Alterações na distribuição sobre um conjunto de 3 classificadores utilizando o algoritmo de Hedge	22
Figura 3.4: Algoritmo de Hedge	23
Figura 3.5: Todos os passos na execução da variante Adaboost	25
Figura 3.6: O rápido decréscimo do erro de treino em função do número de iterações	26
Figura 3.7: Disposição dos padrões a classificar.....	26
Figura 3.8: Execução do Adaboost. a),b),c) e d) classificador final após a realização de 1,3,5 e 10 iterações respectivamente.	27
Figura 4.1: Diferença entre a imagem original e a imagem Pré-processada.	33
Figura 4.2: Slices irrelevantes no reconhecimento da doença do Alzheimer.....	33
Figura 4.3: Localização das máscaras M1 e M2.....	34
Figura 4.4: Matriz das distâncias entre as imagens das classes AD e Normal sobre a máscara M1. a) norma L_2 ; b) norma L_1	35
Figura 4.5: Matriz das distâncias entre as imagens das classes AD e MCI sobre a máscara M2. a) norma L_2 ; b) norma L_1	36
Figura 4.6: Matriz das distâncias entre as imagens das classes MCI e Normal sobre a máscara M1. a) norma L_2 ; b) norma L_1	36
Figura 4.7: Dissemelhança entre imagens da mesma classe.	39
Figura 4.8: Localização das regiões mais distintas na slice 20.	40
Figura 4.9: Localização das regiões mais distintas na slice 35.	41
Figura 4.10: A forma e as dimensões tridimensionais das features de Haar utilizadas na extração de características	42
Figura 4.11: Localização do hipocampo	46
Figura 4.12: As principais 6 regiões determinadas pelo Adaboost.....	47

Lista de Tabelas

Tabela 1.1: A escala e a respectiva classificação do MMSE.....	3
Tabela 2.1: Apresenta-se os resultados dos classificadores sobre as imagens PET e SPECT.13	
Tabela 2.2: Apresenta-se os resultados dos classificadores sobre as imagens MRI.....	14
Tabela 2.3: Apresenta-se os resultados na distinção binária entre as três classes em imagens MRI.	15
Tabela 2.4: Apresenta-se os resultados referenciais.....	15
Tabela 3.1: Decaimento do erro do conjunto de treino.....	28
Tabela 4.1: Apresenta-se as características das imagens PET	32
Tabela 4.2: Apresenta-se os resultados percentuais do método dos k -vizinhos mais próximos.	38
Tabela 4.3: Apresenta-se os resultados percentuais do Adaboost.	44
Tabela 4.4: Apresenta-se os resultados percentuais do Adaboost com redução inicial do vector de características.	45
Tabela 4.5: Comparação dos resultados dos classificadores (%).....	48

Lista de Abreviaturas

AD – Alzheimer’s Disease
PET – Positron Emission Tomography
PPA – Precursor da Proteína Amilóide
MMSE – Mini Mental State Examination
MRI – Magnetic Resonance Imaging
FMRI – Function Magnetic Resonance Imaging
MCI – Mild Cognitive Impairment
FDG – Flurodeoxyglucose
SPECT – Single Photon Emission Computed Tomography
Exact – Exactidão
Sens – Sensibilidade
Esp – Especificidade
CP – Correcto Positivo
CN – Correcto Negativo
FN – Falso Negativo
FP – Falso Positivo
NNC – Nearest Neighbor Classifier
OLC – Optimal Linear Classifier
SVM – Support Vector Machine
CPFIC – Characteristic-Point-based Fuzzy Inference Classification

NMC – Nearest Mean Classifier
PFLD – Pseudo Fisher Linear Distriminant
FLD – Fisher Linear Distriminant
LDA – Linear Discriminant Analysis
QDA – Quadratic Discriminant Analysis
MAP – Máximo a Posteriori
ADNI – Alzheimer’s Disease Neuroimaging Initiative
FWHM – Full Width Half Maximum
LP – Linear Program

1. Introdução

As doenças que causam deterioração progressiva das funções cognitivas, previamente adquiridas, são designadas por demências. Estima-se que actualmente existam 24.3 milhões de indivíduos com demência no mundo [1]. A doença de Alzheimer (AD da designação anglo-saxónica Alzheimer's Disease) é actualmente a demência mais frequente (figura 1.1).

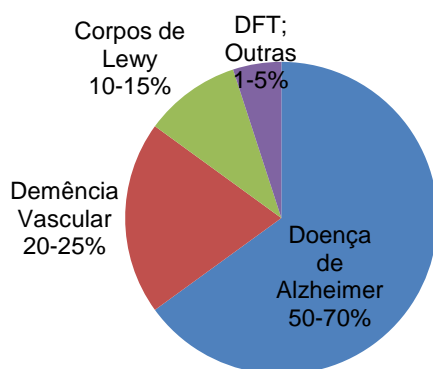


Figura 1.1: Distribuição das demências [1].

A elevada incidência da doença de Alzheimer e a dificuldade na análise visual das imagens *Positron Emission Tomography* (PET), motivaram a utilização dos métodos de reconhecimento de padrões com o objectivo do diagnóstico automático da doença de Alzheimer sobre as imagens PET. A decisão é realizada usando métodos de reconhecimento de padrões, ou seja, métodos capazes de aprender a partir de um conjunto de treino oriundo de decisões anteriores.

1.1. Doença de Alzheimer

No início do século XX, Alois Alzheimer, um médico e investigador alemão, identificou uma doença que provoca alterações degenerativas no tecido nervoso. Ao analisar o cérebro de um doente falecido, com 60 anos, o investigador deparou-se com um caso da doença de Alzheimer, “uma doença invulgar de demência pré-senil”, por ele assim definida. A prevalência da doença de Alzheimer aumenta com idade e, dado o envelhecimento da população, é expectável que a sua incidência aumente (figura 1.2). A causa do surgimento da doença de Alzheimer deve-se à formação de placas senis ou beta-amilóide. As placas senis derivam do precursor da proteína amilóide (PPA), localizada nas terminações nervosas. A sua degradação anormal produz fragmentos peptídeos que se agregam à insolúvel beta-amilóide A4, causando anomalias no funcionamento dos neurónios.

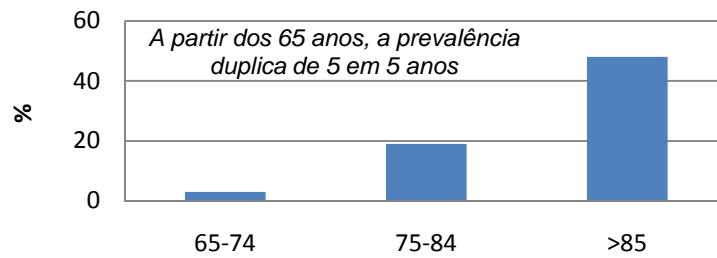


Figura 1.2: A prevalência da doença de Alzheimer aumenta com a idade [1].

Tipicamente o primeiro sinal da doença de Alzheimer corresponde à perda da memória recente, o que é frequentemente observado apenas pelos membros da família. À medida que a perda da memória progride, o doente torna-se desorientado no tempo e depois também no espaço. A depressão aparente nas primeiras fases da doença de Alzheimer origina um estado agitado e sem descanso. Numa fase mais avançada da doença de Alzheimer, há dificuldade no reconhecimento de familiares, perda do relacionamento social e sintomas psiquiátricos, incluído psicoses e alucinações. O início da doença de Alzheimer é insidioso com perda gradual da memória e de outras funções cognitivas (linguagem, orientação, juízo crítico), mas o curso degenerativo é inexorável, culminando na perda total de autonomia.

O lobo temporal, lobo parietal e o hipocampo são as regiões mais propícias ao desenvolvimento da doença de Alzheimer. A doença de Alzheimer afecta inicialmente o lobo temporal¹ e, só depois se desenvolve nas outras regiões especificadas anteriormente. Assim, a análise do lobo temporal tem grande preponderância num diagnóstico precoce da doença de Alzheimer. Na figura 1.3 é possível visualizar as regiões mais sensíveis ao desenvolvimento da doença de Alzheimer.

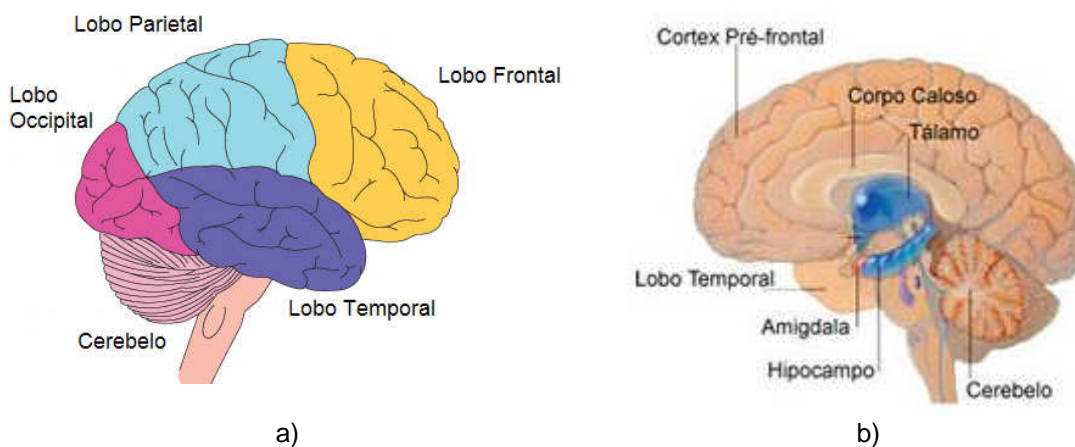


Figura 1.3: Cérebro humano: a) Localização dos lobos [2]; b) Localização do hipocampo [3].

¹ Em Anexo A estão explicadas as funções do córtex cerebral.

Volvidos cem anos, ainda pouco se conhece sobre a patologia que causa a degeneração do tecido nervoso. Não há cura para a doença de Alzheimer, por isso, o principal objectivo do tratamento é minimizar os danos e a propagação da doença. Apenas o diagnóstico precoce pode oferecer um ganho de autonomia e vivacidade nos últimos anos de vida.

1.2. Diagnóstico da Doença de Alzheimer

Quando o doente apresenta os sintomas da doença de Alzheimer, a primeira avaliação diagnóstica resulta dos testes cognitivos. O teste cognitivo mais usual é o *Mini Mental State Examination* (MMSE)², que testa as habilidades de orientação, memória, linguagem, cálculo e capacidade construtiva. Na tabela abaixo é possível visualizar as classificações do MMSE.

Escala	Classificação
24-30	Normal
20-23	Mild Cognitive Impairment (MCI)
10-19	Alzheimer
0-9	Alzheimer terminal

Tabela 1.1: A escala e a respectiva classificação do MMSE.

O MCI é um estado de transição entre o estado Normal e o estado de Alzheimer. O doente já se encontra numa situação de Alzheimer precoce, contudo consegue idealizar raciocínios, participar em actividades sociais e viver de forma autónoma nos restantes anos, caso a doença de Alzheimer não progrida. Se o MCI progredir, os sintomas do Alzheimer evidenciam-se.

As imagens PET e *Magnetic Resonance Imaging* (MRI) indicam as alterações funcionais e estruturais, respectivamente, no cérebro humano. A doença de Alzheimer provoca a redução do fluxo sanguíneo e a diminuição do volume nas regiões sensíveis à sua propagação, assim juntamente com o MMSE, o exame PET e MRI são fundamentais para diagnosticarem a doença de Alzheimer.

1.3. Estudos Previamente Realizados no Diagnóstico Automático da Doença de Alzheimer

A dificuldade na análise visual de imagens MRI e principalmente de imagens PET, no diagnóstico da doença de Alzheimer, originou a utilização dos métodos de reconhecimento de padrões da doença de Alzheimer em imagens MRI e PET como ferramentas de diagnóstico automático da doença de Alzheimer.

² No anexo B está um exemplo de um teste MMSE.

Com o intuito de obter um diagnóstico automático rigoroso da doença de Alzheimer usando as imagens MRI e PET, foram realizados vários estudos com diferentes implementações. A extracção de características e o desempenho dos classificadores foram dois pontos fulcrais da análise em cada estudo. Muitas estratégias foram aplicadas na extracção de características, sempre com o propósito de extrair as características mais discriminativas capazes de identificar correctamente os padrões da doença de Alzheimer.

Na grande maioria dos estudos, apenas se pretendeu realizar a distinção binária entre as classes AD e Normal sobre as imagens MRI e PET. A classe MCI raramente foi implementada na discriminação com as outras classes. Os exames MRI e PET oferecem diferentes informações do cérebro humano, contudo os desempenhos dos classificadores na distinção entre as classes AD e Normal são semelhantes.

1.4. Objectivo da Dissertação

Os métodos dos k -vizinhos mais próximos e de Boosting são utilizados na distinção binária entre as classes AD e Normal, AD e MCI e, finalmente entre as classes MCI e Normal. O método dos k -vizinhos mais próximos é bastante simples e intuitivo. Por outro lado, o método de Boosting é mais complexo e tem a capacidade de melhorar a exactidão de qualquer classificador, especialmente dos classificadores simples. Os métodos de reconhecimento de padrões utilizados na identificação da doença de Alzheimer nas imagens PET estão sujeitos ao problema da “maldição da dimensionalidade”. A dificuldade de aprendizagem sobre um número reduzido de padrões com dimensão elevada designa-se por “maldição da dimensionalidade”. O volume do cérebro numa imagem PET contém milhares de *voxels*, que na grande maioria não possuem informações relevantes sobre a doença de Alzheimer e que apenas acrescentam ambiguidade. A ideia principal do método de Boosting consiste na combinação de classificadores, de forma a superar as suas limitações individuais e beneficiar do bom desempenho de cada classificador em situações específicas. Desta forma, o método de Boosting consegue resistir bem à elevada dimensionalidade das imagens PET.

Devido à dificuldade na análise visual e interpretação das imagens PET, o objectivo da utilização dos métodos dos k -vizinhos mais próximos e de Boosting é torná-los numa ferramenta útil e eficaz de diagnóstico automático da doença de Alzheimer. Outro objectivo consiste na verificação do desempenho destes métodos na distinção binária entre as classes AD e MCI, e principalmente entre as classes MCI e Normal. De salientar, a preponderância da distinção entre as classes MCI e Normal num diagnóstico precoce da doença de Alzheimer.

O diagnóstico precoce da doença de Alzheimer permite travar sua progressão e simultaneamente oferecer melhores condições e autonomia para enfrentar os restantes anos de vida.

1.5. Contribuições da Dissertação

Esta dissertação apresenta um estudo preciso no âmbito do diagnóstico automático da doença de Alzheimer. Utiliza um conjunto vasto de 195 imagens PET, igualmente divididas pelas classes AD, MCI e Normal, o que proporciona grande precisão no estudo. Por outro lado, a utilização da validação cruzada na avaliação dos métodos dos k -vizinhos mais próximos e de Boosting, garante resultados correctos no desempenho destes métodos.

De referir, a implementação da classe MCI na discriminação entre as classes AD e MCI e, também entre as classes MCI e Normal, raramente estudadas. A discriminação entre as classes MCI e Normal permite diagnosticar a doença de Alzheimer num estado inicial, fulcral no tratamento e retardamento da evolução da doença. A discriminação entre as classes AD e MCI concede uma medida sobre o estado da evolução da doença de Alzheimer. Ambas as discriminações são complicadas devido à semelhança das imagens PET, contudo facultam informações importantes.

Este estudo aplica o método de Boosting na identificação na doença de Alzheimer sobre as imagens PET. É um método ainda pouco testado nas imagens PET [4]³, embora já tenha sido usado em imagens MRI [5,6]. Este estudo averigua se as virtudes inerentes ao método de Boosting proporcionam um bom desempenho na distinção binária entre as classes.

1.6. Organização da Dissertação

A presente dissertação organiza-se em cinco capítulos e 2 anexos. O primeiro capítulo, Introdução, descreve o problema abordado: doença de Alzheimer. Por outro lado, indica o caminho na resolução do problema através do diagnóstico automático da doença de Alzheimer, e finalmente mostra a motivação no estudo do problema.

No segundo capítulo descrevem-se os estudos previamente realizados sobre o diagnóstico automático da doença de Alzheimer. Sucintamente explica-se a estrutura de um sistema de reconhecimento de padrões e os métodos de avaliação dos classificadores. O diagnóstico automático da doença de Alzheimer pode ocorrer sobre as imagens PET e MRI. Discutem-se as diferenças entre as imagens PET e MRI. Fundamentalmente neste segundo capítulo apresenta-se o desempenho dos classificadores descritos na literatura, na distinção binária entre as classes AD e Normal sobre as imagens PET e MRI.

No terceiro capítulo explicam-se pormenorizadamente os métodos dos k -vizinhos mais próximos e de Boosting. Refere-se as vantagens e desvantagens de ambos os métodos e, justificam-se as suas utilizações. Antes de realizar a classificação das imagens PET e MRI, é necessário extrair as características fundamentais que identificam a doença de Alzheimer.

³ Este estudo foi publicado 2009.

Debruça-se sobre a enorme importância da extração e selecção de características na execução da classificação.

No quarto capítulo apresentam-se os resultados dos métodos dos k -vizinhos mais próximos e de Boosting sobre as imagens PET. Analisa-se os resultados e realiza-se uma comparação entre os resultados obtidos. O quinto capítulo apresenta-se uma sucinta conclusão da dissertação e apresentam-se propostas para trabalhos futuros.

Ainda fazem parte da dissertação dois anexos de A a B. No primeiro anexo explica-se a função do hipocampo, do lobo temporal, parietal, occipital e frontal. No anexo B fornece-se a estrutura base dos testes cognitivos MMSE, através de um teste tipo.

2. Revisão dos Classificadores Utilizados na Identificação da Doença de Alzheimer em Imagens PET e MRI

Neste capítulo apresenta-se uma introdução sucinta ao reconhecimento de padrões e explica-se a estrutura de um sistema de reconhecimento de padrões. O sistema de reconhecimento de padrões classifica uma imagem PET ou MRI através de um *conjunto de características* discriminativas que identificam a doença de Alzheimer. Na primeira fase realiza-se uma aprendizagem do conjunto de características através do treino do classificador. O treino do classificador gera uma estrutura capaz de, numa segunda fase, decidir e atribuir uma classe à imagem PET ou MRI. Por outro lado, neste capítulo apresenta-se o desempenho de vários classificadores discutidos na literatura, na identificação da doença de Alzheimer sobre as imagens PET e MRI.

2.1. Reconhecimento de Padrões

2.1.1. Extração de Características e o Classificador

A estrutura de um sistema de reconhecimento de padrões é essencialmente constituída por dois blocos, visíveis na figura 2.1 [7]. O primeiro bloco realiza a função de *extração de características*. Este primeiro bloco extrair as características fundamentais e mais discriminativas capazes de identificar correctamente um determinado padrão. A perfeita extração de características é preponderante na precisão do *classificador*. O classificador determina e atribui a classe que melhor se adapta às características do padrão.

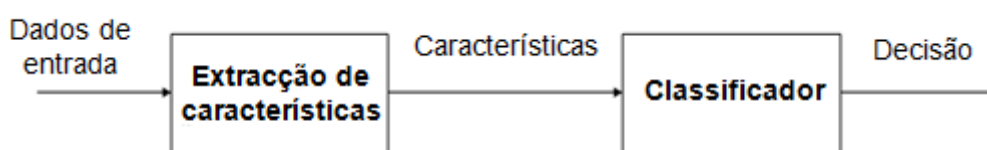


Figura 2.1: Estrutura de um sistema de reconhecimento de padrões [7].

O resultado da extração de características representa-se normalmente por um *vector de características*. O treino do classificador pode ser realizado em dois modos: O *modo supervisionado* e *não supervisionado*. Na aprendizagem supervisionada, o classificador adapta-se em função do conjunto de padrões de entrada, *padrões de treino*, para os quais a classe desejada de saída é conhecida. Assim, na aprendizagem supervisionada, a classe que gera cada padrão de treino é conhecida *a priori*. Na aprendizagem não supervisionada, os

padrões de treino não se encontram rotulados, o que implica que o classificador se adapta aos padrões procurando encontrar as classes naturais.

A capacidade de generalização do classificador é o objectivo vital na realização do seu treino. É importante que o classificador possuía uma estrutura capaz de classificar correctamente os padrões não presentes no treino, *padrões de teste*. Na eventualidade do classificador se adaptar demasiado bem aos padrões de treino, pode originar a sobreaprendizagem dos padrões de treino e a fraca generalização.

2.1.2. Avaliação do Desempenho Utilizando Validação Cruzada

Após a realização do treino do classificador é necessário avaliar o seu desempenho para um ou vários conjuntos de padrões de teste. A forma mais simples de avaliar o classificador é dividir o conjunto de padrões num conjunto de padrões de treino e num conjunto de padrões de teste. Quando o conjunto total de padrões é reduzido, a utilização de apenas um único conjunto de padrões de teste pode não ser suficiente para garantir uma avaliação correcta.

Nessa situação, a validação cruzada é uma alternativa para avaliar o classificador. A validação cruzada consiste em dividir o conjunto de padrões em subconjuntos com o mesmo número de padrões. Aleatoriamente, um subconjunto de padrões é identificado como de teste e os restantes subconjuntos de padrões como de treino. Após a avaliação do classificador para o respectivo subconjunto de padrões de teste, ocorre uma rotação na atribuição dos subconjuntos, para que outro subconjunto de padrões seja o de teste. Este procedimento é repetido até que seja atribuída a condição de subconjunto de padrões de teste a todos os subconjuntos de padrões. Apesar do número reduzido de padrões, desta forma é possível obter um resultado rigoroso na avaliação do desempenho do classificador.

A desvantagem deste método é a elevada sobrecarga computacional, pois é necessário realizar o treino do classificador o número de vezes igual ao número de divisões do conjunto de padrões. Uma variante do método validação cruzada é a utilização de apenas um único padrão na realização do teste do classificador. Esta variante é conhecida como *leave-one-out*, sendo utilizada particularmente em quantidades precárias de padrões.

2.2. Reconhecimento de Padrões de Alzheimer nas Imagens PET e MRI

Nos últimos anos, efectuaram-se alguns estudos com o objectivo principal de especificar padrões nas imagens PET e MRI do cérebro humano, capazes de identificar a existência de determinadas demências. As imagens PET e MRI permitem visualizar a actividade e a estrutura tridimensional do cérebro, de uma forma não invasiva (visualizar o interior do cérebro sem recorrer à cirurgia). Com a realização continua de exames PET ou MRI, de seis em seis meses, sobre uma população de pacientes com demências é possível visualizar nas imagens respectivas a localização, a evolução e a trajectória da demência no cérebro do paciente.

2.2.1. Imagens PET

A capacidade de detectar alterações funcionais e metabólicas no cérebro humano, mesmo antes das alterações anatómicas se manifestarem, é a característica fundamental das imagens de Medicina Nuclear. Para tal, utilizam-se quantidades mínimas de substâncias radioactivas que acompanham, sem interferir, os processos metabólicos. O radiofármaco Flurodeoxyglucose (FDG) é injectado no paciente. O FDG é composto por uma molécula de glicose, combustível básico de energia das células do cérebro, ligado artificialmente a um átomo de flúor radioactivo. As células do cérebro com maior actividade necessitam de absorver maiores quantidades de FDG. O átomo de flúor na molécula FDG sofre um decaimento radioactivo, emitindo um positrão (mesma massa que um electrão, mas com carga simétrica). Quando o positrão colide com um electrão ocorre uma aniquilação, que liberta energia em forma de dois raios gama, orientados em direcções opostas. Daí advém o nome Positron Emission Tomography (PET). As regiões do cérebro onde ocorrem cada impulso de radiação são computacionalmente reconstruídas. As regiões que absorvem mais FDG emitem maiores quantidades de radiação, isto é, são regiões com maior metabolismo cerebral ou circulação sanguínea, representadas nas imagens PET pelos voxels mais intensos (os voxels de cor branca representam maior intensidade e os voxels de cor preta menor intensidade). Através da pouca intensidade em determinadas regiões, devido a fraca actividade sanguínea, é possível diagnosticar a doença de Alzheimer no paciente. Na figura 2.2 é possível visualizar uma imagem PET pertencente à classe AD e à classe Normal em dois cortes axiais.

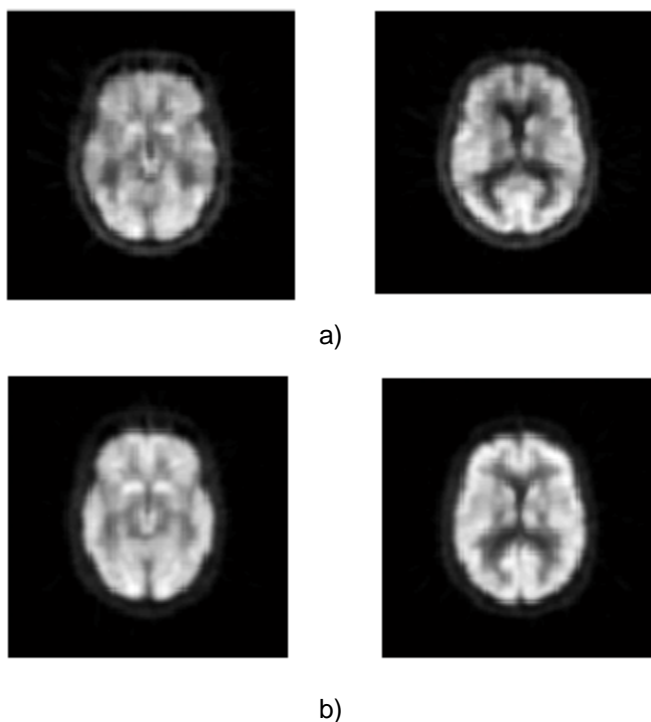


Figura 2.2: Imagem PET em corte axial: a) classe AD; b) classe Normal.

O diagnóstico precoce da doença de Alzheimer é essencial na manutenção da vivacidade do doente. As imagens PET possuem elevada sensibilidade e enorme capacidade de identificar as alterações provocadas pelas demências [8], incluindo a doença de Alzheimer. Através da análise das imagens PET, é possível atingir maior exactidão no diagnóstico da doença de Alzheimer num estado inicial do que nas imagens MRI [8].

As imagens Single Photon Emission Computed Tomography (SPECT), à semelhança das imagens PET, oferecem a informação funcional sobre as regiões do cérebro. As imagens SPECT também fazem parte das imagens de Medicina Nuclear, contudo na realização do exame SPECT não existe emissões de positrões, mas sim de fotões.

2.2.2. Imagens MRI

A imagem MRI é uma técnica de imagiologia médica utilizada para visualizar a estrutura interna do corpo. As imagens MRI são obtidas através das propriedades magnéticas do núcleo do hidrogénio e, das suas interações com os campos magnéticos externos e com as ondas de rádio. Durante a realização de um exame MRI, os fortes electromagnetes criam um campo magnético que alinham os protões dos átomos de hidrogénio na direcção do campo. A emissão das ondas rádio provoca o desalinhamento dos protões relativamente ao campo magnético. Quando o impulso rádio é desligado, os protões realinham-se com o campo magnético inicial. Esta mudança de alinhamento emite radiações hiperfinas que produzem as respectivas imagens MRI. Os protões dos tecidos deteriorados realinham-se ao campo magnético inicial em instantes distintos, criando contrastes nas imagens MRI.

Na imagem MRI cerebral representa-se a estrutura anatómica do cérebro humano. A doença de Alzheimer provoca alterações na integridade e densidade dos tecidos do cérebro. Estas alterações levam ao atrofio do cérebro humano. O atrofio do cérebro significa a redução total do volume do cérebro, pelo decréscimo da densidade das sinapses e da redução da "massa cinzenta". Nas regiões mais propícias ao desenvolvimento do Alzheimer, a redução do volume é mais acentuada. Os pacientes com Alzheimer sofrem uma redução de 25% no volume do hipocampo, relativamente aos pacientes Normais. Por outro lado, os pacientes no estado MCI têm uma redução da ordem de 11% [9]. A análise das imagens MRI evidencia as alterações no volume do cérebro, sendo assim, o exame MRI é uma ferramenta de diagnóstico da doença de Alzheimer. A figura 2.3 mostra uma imagem MRI pertencente à classe AD e à classe Normal em corte axial.

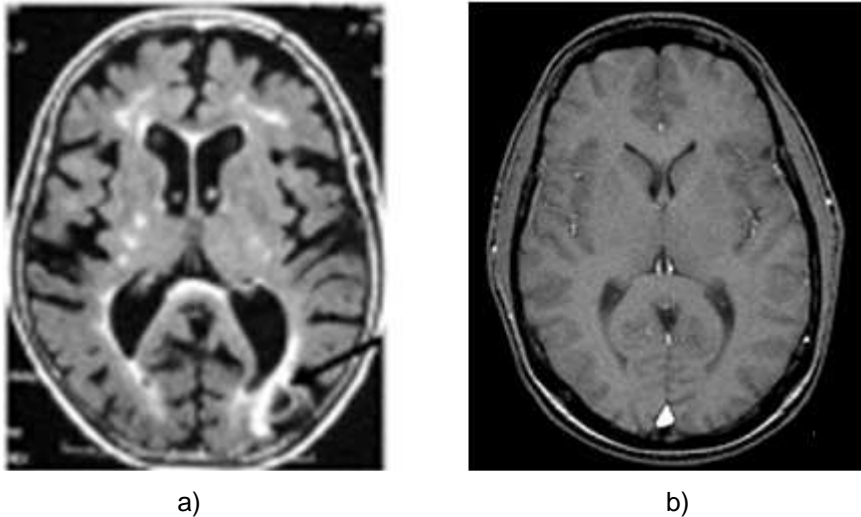


Figura 2.3: Imagem MRI em corte axial: a) classe AD; b) classe Normal [10].

2.3. Extração de Características das Imagens PET e MRI

As imagens PET e MRI contêm milhares de voxels, cada um deles com um valor de intensidade associado. A intensidade dos voxels representa o grau de actividade ou densidade do tecido no cérebro humano. Os valores das intensidades dos voxels são utilizados na determinação de características sobre as imagens PET e MRI. As características extraídas são colocadas num vector de características, que posteriormente será preponderante na decisão do classificador. Assim, é essencial que as características inseridas no vector de características sejam as mais discriminativas, para treinar o classificador correctamente. Nos estudos já realizados com as imagens PET, optou-se por várias estratégias na extração de características, mas em todos os estudos utilizou-se como base as intensidades dos voxels. Nos estudos com as imagens MRI, utilizou-se as intensidades dos voxels e os dados volumétricos. Em seguida apresentam-se a maioria dos vectores de características usados no treino do classificador:

- Todas as intensidades dos voxels do cérebro [15,18];
- Apenas as intensidades dos voxels de alguns lobos do cérebro (parietal e temporal) [13];
- Conjunto reduzido de intensidades dos voxels (inferior a 100 voxels) nas regiões propícias ao desenvolvimento da doença de Alzheimer [11,13,14,6,17];
- Operações aritméticas entre conjuntos de intensidades dos voxels [12].

De salientar que os melhores resultados surgem nos vectores de características reduzidos, isto significa, que existiu uma prévia selecção das características extraídas. A figura 2.4 representa 7 áreas extraídas de uma imagem PET, resultando num vector de características com 253 voxels.

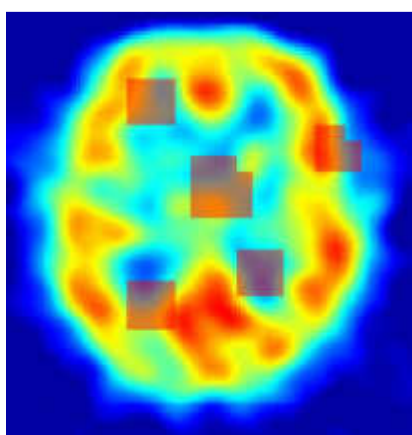


Figura 2.4: As localizações das sete áreas utilizadas no treino do classificador [11].

2.4. Resultados dos Classificadores Utilizados na Detecção da Doença de Alzheimer sobre as Imagens PET e MRI

A dificuldade em diagnosticar a doença de Alzheimer através da análise visual e interpretação das imagens PET e MRI, provocou a necessidade da utilização de métodos de reconhecimento de padrões para classificar a imagem na classe AD ou classe Normal. Bastantes classificadores foram testados com a finalidade de verificar a sua exactidão na distinção entre as classes AD e Normal. Na tabela 2.1 é possível visualizar a Exactidão (Exact), Sensibilidade (Sens), Especificidade (Esp) de vários classificadores sobre as imagens PET e SPECT, as quantidades de imagens utilizadas e os métodos de avaliação dos classificadores. Os parâmetros de avaliação do classificador foram obtidos a partir das seguintes expressões:

$$\text{Sensibilidade} = \frac{CP}{CP+FN} \quad (2.1)$$

$$\text{Especificidade} = \frac{CN}{CN+FP} \quad (2.2)$$

$$\text{Exactidão} = \frac{CP+CN}{CP+CN+FP+FN} \quad (2.3)$$

onde o correcto positivo (CP) é o numero de imagens AD correctamente classificadas; correcto negativo (CN) é a quantidade de imagens Normais correctamente classificadas; falso negativo (FN) é o número de imagens AD classificadas como Normais; falso positivo (FP) é a quantidade de imagens Normais classificadas como AD. Logo a sensibilidade é a percentagem de imagens AD bem classificadas e a especificidade é a percentagem de imagens Normais correctamente classificadas.

Classificador	Exact. AD/N(%)	Sens. AD/N(%)	Esp. AD/N(%)	Nº imagens AD	Nº imagens N	Avaliação do classificador
NNC [12]	85.1	-	-	58 (22M/36F)	49 (22M/27F)	Validação Cruzada
OLC [12]	90.8	85.9	96.8			
NNC [13]	83.0	-	-	58 (22M/36F)	49 (22M/27F)	Validação Cruzada
SVM [13]	85.4	-	-			
CPFIC [13]	84.7	-	-			
NMC [14]	84.8	79.3	88.0	29	50	Leave-one-out
PFLD [14]	89.9	82.8	94.0			
Retropropagação [15]	92.3	91.0	93.0	22	30	26 Treino 26 Teste
SVM [11]	87.6	84.4	90.9	92	31	90 Treino 33 Teste
FLD [11]	84.7	82.0	87.5			
LP boosting [4]	84.0	84.0	82.0	149 (88M/61F)		Validação Cruzada

Nearest Neighbor Classifier (NNC); Optimal Linear Classifier (OLC); Support Vector Machine (SVM); Characteristic-Point-based Fuzzy Inference Classification (CPFIC); Nearest Mean Classifier (NMC); Pseudo Fisher Linear Discriminant (PFLD); Fisher Linear Discriminant (FLD); Linear Program (LP).

Tabela 2.1: Apresenta-se os resultados dos classificadores sobre as imagens PET e SPECT.

A maioria dos classificadores possui uma exactidão na discriminação entre a classe AD e a classe Normal próxima dos 85%, em que o classificador de Retropropagação atingiu a exactidão de 92.3% e o NNC (Nearest Neighbor Classifier) atingiu a exactidão de 83.0%, no limite superior e inferior, respectivamente, da exactidão na discriminação entre as classes AD e Normal. De forma a verificar o grau de precisão inerente ao classificador é necessário analisar a quantidade de imagens utilizadas e o método de avaliação do classificador. O principal objectivo no treino do classificador consiste na capacidade de generalização para conjuntos de imagens disjuntas ao conjunto de imagens de treino. O estudo que utilizou o classificador de Retropropagação atingiu a exactidão mais alta, contudo nesse estudo utilizou-se um número reduzido de imagens e a avaliação da estrutura do classificador realizou-se sobre apenas um conjunto de imagens de teste. Sendo assim, apesar de possuir uma exactidão elevada na

distinção entre as classes AD e Normal, não é garantido que o classificador generalize igualmente bem para outros conjuntos de imagens de teste. Todos os classificadores atingiram um valor de especificidade superior ao valor de sensibilidade. Isto significa que os classificadores têm maior capacidade de reconhecer correctamente as imagens Normais. A tabela 2.2 representa o desempenho dos classificadores no diagnóstico automático da doença de Alzheimer sobre as imagens MRI. Também mostra o respectivo conjunto de imagens MRI utilizadas e o método de avaliação do classificador.

Classificador	Exac. AD/N (%)	Sens. AD/N (%)	Esp. AD/N (%)	Nº imagens AD	Nº imagens N	Avaliação do classificador
LDA [16]	84.0	82.0	85.0	75	75	Leave-one-out
QDA [16]	83.3	84.0	83.0			
SVM [16]	92.0	91.0	93.0			
AdaboostSVM [6]	86.0	80.0	92.0	49F	49F	Validação Cruzada
LDA [17]	91.0	79.0	100	24	40	30 Treino 34 Teste
Retro-propagação [18]	83.7	82.6	84.7	25 (5M/20F)	27 (17M/10F)	O conjunto de dados é dividido em dois conjuntos disjuntos de teste e treino

Linear Discriminant Analysis (LDA); Quadratic Discriminant Analysis (QDA).

Tabela 2.2: Apresenta-se os resultados dos classificadores sobre as imagens MRI.

Apesar das técnicas PET e MRI revelarem diferentes informações do cérebro humano, os resultados da exactidão na distinção entre as classes AD e Normal foram bastante semelhantes. De referir, tal como aconteceu nas imagens PET, a maior facilidade em reconhecer as imagens Normais.

De salientar, o classificador SVM utilizado maioritariamente nos vários estudos realizados, atingindo sempre rácios elevados de exactidão. O objectivo máximo do SVM é obter um plano que separe as classes distintas. Em determinados casos há vários planos possíveis para separar as classes diferentes, mas o plano que maximiza a distância entre padrões inseridos na fronteira de separação (*vectores de suporte*) das classes distintas, é o escolhido. Na maioria das situações as classes não são linearmente separáveis. Através das funções *kernel* é possível aumentar a dimensionalidade dos padrões de entrada para o *espaço de características*, onde se torna possível separar as classes com um hiperplano. O classificador SVM, referido na tabela acima, atingiu uma exactidão de 92%, sobre um rigoroso método de avaliação da estrutura do classificador.

De realçar um estudo raro que considerou a classe MCI sobre as imagens MRI e, assim, tornou possível realizar a distinção entre as classes AD e MCI e, MCI e Normal [18]. O conjunto total de imagens foi constituído por 86 imagens MRI, nas quais 25 imagens pertenciam à classe AD, 27 imagens eram da classe Normal e finalmente, 16 imagens foram inseridas na classe MCI. Na tabela 2.3 é possível visualizar a exactidão, sensibilidade e especificidade do classificador FLD na distinção binária entre as três classes.

Classificador: FLD [18]	AD/N (%)	AD/MCI (%)	MCI/N (%)
Exactidão	82.8	62.5	80.7
Sensibilidade	80.4	58.9	81.3
Especificidade	85.0	68.0	80.3

Tabela 2.3: Apresenta-se os resultados na distinção binária entre as três classes em imagens MRI.

A classe MCI é um estado inicial da doença de Alzheimer e, por isso, fez diminuir consideravelmente o rácio de exactidão na distinção entre as classes AD e MCI. O classificador nesta discriminação está a separar imagens de pacientes com Alzheimer, mas em diferentes estados de evolução. Por outro lado, neste estudo a exactidão do classificador na separação entre as classes AD e Normal e as classes MCI e Normal foi semelhante. Nesses casos, o classificador está a distinguir estados distintos, Alzheimer ou Normal. Neste estudo utilizou-se um número reduzido de imagens e o método de avaliação do classificador realizou-se sobre um único conjunto de imagens de teste. Assim, a avaliação do desempenho deste classificador foi pouco rigorosa. Neste capítulo, foram relatados os resultados obtidos por vários classificadores, em diferentes condições de treino e teste. São resultados que devem ser tomados em consideração no estudo a realizar nos capítulos posteriores. Especialmente há dois estudos, que serão uma referência e uma meta nos resultados a obter nesta dissertação, pela similaridade na quantidade de imagens e no método de avaliação utilizado (tabela 2.4).

Classificador	Imagem	Exact. AD/N(%)	Sens. AD/N(%)	Esp. AD/N(%)	Nº imagens AD	Nº imagens N	Avaliação do classificador
OLC [12]	PET	90.84	85.86	96.79	58 (22M/36F)	49 (22M/27F)	Validação Cruzada
AdaboostSVM [6]	MRI	86.00	80.00	92.00	49F	49F	Validação Cruzada

Tabela 2.4: Apresenta-se os resultados referenciais.

3. Diagnóstico Automático da Doença de Alzheimer

A doença de Alzheimer é uma realidade obscura. Volvido um século, ainda pouco se conhece sobre a patologia que progressivamente e de forma irreversível destrói o tecido nervoso do cérebro provocando a completa ausência de autonomia. A difícil análise visual das imagens PET provocou a necessidade de implementar vários métodos de Reconhecimento de Padrões como possíveis ferramentas ao diagnóstico automático da doença de Alzheimer. Neste capítulo, apresentam-se os métodos dos k -vizinhos mais próximos e de Boosting, utilizados nesta dissertação no diagnóstico automático da doença de Alzheimer sobre as imagens PET. Ambos os métodos foram seleccionados dentro de um conjunto enorme de outros possíveis métodos de classificação.

O método dos k -vizinhos mais próximos foi previamente utilizado no diagnóstico automático da doença de Alzheimer, atingindo bons resultados [12,13]. O método dos k -vizinhos mais próximos é bastante prático e intuitivo. Neste método, a classificação consiste na atribuição da classe maioritária dos k padrões de treino mais próximos ao padrão desconhecido x . A execução simples e os resultados do desempenho deste método serviram de plataforma e referência na comparação com um método mais complexo.

O método de Boosting é um método complexo capaz de melhorar a exactidão de qualquer classificador, em especial dos classificadores simples. Escassas são as referências da utilização do método de Boosting sobre as imagens PET [4], mas sobre as imagens MRI atingiu-se bons desempenhos [5,6]. A grande capacidade do método de Boosting na adaptação ao problema da “maldição da dimensionalidade” foi preponderante na sua escolha, de forma a combater a dimensionalidade elevada das imagens PET. A utilização do método de Boosting revolucionou e permitiu realizar rápidas detecções de faces [19]. A introdução de uma nova representação da imagem, de nome “Imagem Integral”, de uma variante do boosting, chamada Adaboost, e finalmente, da combinação dos classificadores em “cascata”, permitiu obter um rápido processamento das imagens [19]. De salientar, a preponderância da variante Adaboost na selecção das características cruciais e no treino do classificador.

3.1. Extração de Características

A primeira fase do reconhecimento de padrões é a extração de características, que *a posteriori* serão seleccionadas e utilizadas no treino do classificador. Esta fase inicial é crucial para o bom desempenho do reconhecedor de padrões, pois a decisão depende totalmente dos dados inseridos no vector de características. As imagens PET do cérebro humano são constituídas por milhares de voxels, que contêm informações sobre a actividade cerebral. Esses milhares de voxels possuem informações dispersantes sobre vários aspectos da funcionalidade do cérebro. Contudo a doença de Alzheimer propaga-se em regiões específicas

do cérebro humano, assim, é vantajoso incluir somente os dados emergentes dessas regiões no vector de características. Por outro lado, há a possibilidade de utilizar algoritmos de selecção de características. Num conjunto numeroso de características, o algoritmo selecciona as características essenciais para obter uma decisão rigorosa e precisa. Um exemplo de um algoritmo de selecção de características é o Adaboost, que irá ser explicado nas próximas secções. O valor inserido no vector de características pode conter o valor exacto da intensidade do voxel ou o valor de operações sobre quadriláteros de voxels.

3.1.1. Features de Haar

As features de Haar, como é possível visualizar na figura 3.1, obtiveram excelentes desempenhos na rápida e eficaz detecção de faces. As features de Haar possibilitam conter informações sobre uma área ou volume do cérebro, em vez de um valor de intensidade isolado.

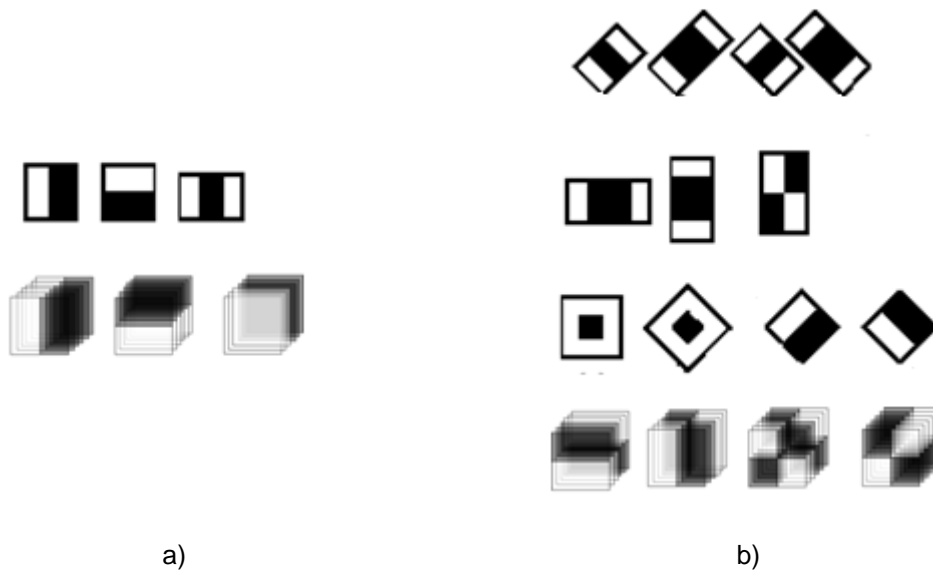


Figura 3.1: Features de Haar a) simples; b) extensões [19].

O valor final obtido é a subtracção do somatório das intensidades dos voxels na região branca com o somatório das intensidades dos voxels na região preta. O valor final obtido pelas features de Haar representa a distinção entre duas regiões distintas. O modelo (dimensão e forma) da feature de Haar ($T_k(x, y, z)$) percorre todas as posições da imagem ($I(x, y, z)$) PET, de forma a extrair as características. Na equação 3.1 representa-se o cálculo das características sobre um modelo tridimensional da feature de Haar.

$$f_k(x_0, y_0, z_0) = \iiint I(x, y, z) \cdot T_k(x - x_0, y - y_0, z - z_0) dx dy dz \quad (3.1)$$

3.2. Métodos dos k -Vizinhos Mais Próximos e de Boosting

3.2.1. Métodos dos k -Vizinhos Mais Próximos

O método dos k -vizinhos mais próximos é um sistema de aprendizagem supervisionada. Este método é um caso particular da classificação de *máximo a posteriori* (MAP), que consiste na minimização da probabilidade de erro do classificador. Assim, o padrão x é associado à classe (\hat{c}) mais provável inserida num conjunto de classes $\Omega = \{c_1, \dots, c_p\}$ definida por:

$$\hat{c} = \arg \max_{c \in \Omega} P(c|x), \quad (3.2)$$

em que $P(c|x)$ é a probabilidade de ocorrer a classe c , dado que se observou o padrão x , designada por *probabilidade a posteriori*. O cálculo da *probabilidade a posteriori* realiza-se através da lei de Bayes:

$$P(c|x) = \frac{p(x|c)P(c)}{p(x)} \quad (3.3)$$

onde $P(c)$ é a *probabilidade a priori*, facilmente estimada a partir do conjunto de treino. O factor de normalização, $p(x)$, não influencia a decisão. Finalmente, $p(x|c)$ é a distribuição dos dados gerados pela classe c , difícil de estimar.

A estimação de $p(x|c)$ pode ser realizada por métodos paramétricos ou não-paramétricos. Os métodos paramétricos são aplicados quando se conhece *a priori* a família paramétrica de distribuições $p(x|\theta)$ que gerou o padrão x , em que θ significa um vector de parâmetros desconhecidos. A distribuição normal é amplamente utilizada, devido às suas inerentes propriedades estatísticas. Em contraste, nos métodos não-paramétricos, a distribuição que gerou o conjunto de padrões é desconhecida. No entanto, é possível determinar a estimativa da distribuição a partir do conjunto de padrões.

O primeiro passo na realização do método dos k -vizinhos mais próximos é escolher o número dos k padrões treino. O resultado da classificação depende preponderantemente do valor escolhido para k . Sobre a localização do padrão desconhecido x coloca-se o centro de uma esfera, de raio variável, com o intuito de conter precisamente os k padrões de treino.

A estimação da densidade no local onde se situa o padrão desconhecido x é obtida recorrendo às seguintes demonstrações [20]:

- A probabilidade P do vector x estar contido na região V é dada pela seguinte equação:

$$P = \int_V p(x) dx, \quad (3.4)$$

se $p(x)$ for uma função continua e a região V tão pequena que $p(x)$ não varia, é possível realizar a seguinte simplificação:

$$P = \int_V p(x) dx \approx p(x)V \quad (3.5)$$

- Se existir N padrões, a probabilidade de k dos N padrões fazerem parte de V é:

$$P_K = \binom{N}{K} P^K (1 - P)^{N-K}, \quad (3.6)$$

em que, $E[k] = NP$.

- A distribuição binomial possui um pico muito acentuado em torno do valor esperado, originando a seguinte simplificação:

$$k \approx E[k] = NP \quad (3.7)$$

- Combinando as equações, resulta:

$$p(x) = \frac{k}{NV} \quad (3.8)$$

O teorema de Bayes é utilizado na obtenção da estimação não-paramétrica de densidade de probabilidade no método dos k -vizinhos mais próximos [20]. Supondo que no total há N padrões, e N_k padrões pertencem a classe C_k , de forma que $\sum_k N_k = N$. O raio da esfera cresce até os k padrões de treino se encontrem completamente dentro da esfera. Supondo que dentro do volume da esfera estão contidos k_k padrões da classe C_k .

- Utilizando a expressão 3.8 é possível obter uma aproximação para a densidade de probabilidade condicional:

$$p(x|C_k) = \frac{K_k}{N_k V} \quad (3.9)$$

- A densidade de probabilidade *a priori* pode ser estimada, resultando:

$$P(C_k) = \frac{N_k}{N} \quad (3.10)$$

➤ Finalmente, utilizando o teorema de Bayes, atinge-se a expressão final:

$$P(C_k|x) = \frac{p(x|C_k)P(C_k)}{p(x)} = \frac{K_k}{K} \quad (3.11)$$

Para minimizar a probabilidade de erro na classificação do padrão desconhecido x , é necessário atribuir a esse padrão x a classe C_k para qual o rácio k_k/k é superior. Sendo assim, a regra do método dos k -vizinhos mais próximos consiste na atribuição directa da classe maioritária dos k padrões de treino mais próximos, ao padrão desconhecido x .

O valor seleccionado para k deve ser um número ímpar, de forma a eliminar a possibilidade de obter classes com a mesma frequência. Por outro lado, tem que existir um compromisso na escolha do valor de k , pois se for muito elevado, poderá apanhar vizinhos maioritariamente pertencentes a outra classe, e se for muito baixo, poderá restringir a classificação, tornando-a pouco fiável.

A métrica mais usual para o cálculo dos k -vizinhos mais próximos é a distância euclidiana, mas não existe qualquer restrição à utilização de outras métricas.

$$d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (3.12)$$

A probabilidade mínima de erro (P_e^*) obtém-se através do classificador MAP, visível na seguinte expressão:

$$P_e^*(x) = 1 - \max_c P(c/x) \quad (3.13)$$

Sendo P_{k-VMP} a probabilidade de erro no método dos k -vizinhos mais próximos, é possível apresentar a seguinte desigualdade [7]:

$$P_e^* \leq P_{k-VMP} \leq P_e^* \left(2 - \frac{n_c}{n_c - 1} P_e^*\right) \quad (3.14)$$

onde n_c é o número de classes. Para um conjunto de treino elevado, a probabilidade de erro do método dos k -vizinhos mais próximos é menor que $2P_e^*$.

3.2.1.1. Exemplo do método dos k -vizinhos mais próximos

Na figura 3.2 apresenta-se um exemplo do método de classificação dos k -vizinhos mais próximos, sobre um conjunto de 15 padrões, em que 7 padrões pertencem à classe azul e os restantes à classe vermelho. O objectivo é atribuir uma classe ao padrão preto através da métrica da distância euclidiana.

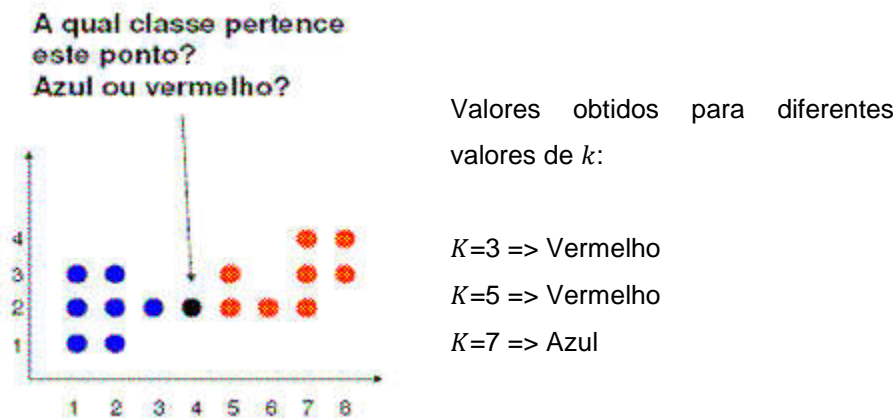


Figura 3.2: Classificação do método dos k -vizinhos mais próximos [21].

Se o número dos k vizinhos for igual a 3, a classe atribuída ao padrão preto é vermelho. O centro da esfera é colocado na localização do padrão a classificar e, o seu raio cresce até conter três padrões de treino dentro da esfera. A classe maioritária correspondente a esses três padrões de treino é a classe vermelha. Sendo assim, o padrão preto é classificado na classe vermelha. É possível verificar que a classificação do padrão preto, tal como já foi referido, permuta de acordo com o valor de k .

O método dos k -vizinhos mais próximos revela ser simples e intuitivo, mas tem a desvantagem dos resultados derivarem de estimações da densidade de probabilidade sensíveis as características irrelevantes e de ser computacionalmente lento para valores elevados de k e padrões de treino numerosos.

3.2.2. Método de Boosting

O método de Boosting também é um método de aprendizagem supervisionada. A ideia principal do método de Boosting consiste no treino de vários *classificadores fracos* de forma iterativa. Após o término do ciclo de iterações, realiza-se a combinação de todos os classificadores fracos previamente treinados, de forma a obter um único *classificador forte*. Em cada iteração, os classificadores fracos tomam decisões consoante as diferentes distribuições dos pesos sobre os padrões de treino. A combinação final dos classificadores fracos tem a capacidade de superar as limitações individuais de cada classificador fraco.

O método Boosting pode ser executado em dois modos [22]: *Boosting by sampling* e *Boosting by filtering*. No primeiro modo extrai-se as características mais discriminativas ao longo das iterações, sobre todos os padrões treino. O objectivo é extrair diversificadas características capazes de englobar todos os padrões de treino. Em algumas aplicações é complicado guardar todos os dados do conjunto de treino na memória do processador. Sendo assim, de forma a eliminar este problema, o *Boosting by filtering* realiza uma selecção dos padrões de treino que serão utilizados pelos classificadores fracos. De forma geral, o método

de Boosting é um excelente método capaz de melhorar a precisão de qualquer classificador, especialmente dos classificadores fracos. Há várias variantes do Boosting, mas sobressaem-se duas variantes utilizadas em muitas situações [23]:

- Adaboost (classificação Binária);
- Adaboost.M1 e Adaboost.M2 (classificação em multi-classes);

Antes de explicar o funcionamento destas duas variantes do Boosting, apresenta-se o primordial algoritmo Hedge, que posteriormente origina o Boosting.

3.2.2.1. Algoritmo Hedge

O método de Boosting foi inspirado pelo algoritmo de aprendizagem em tempo real, designado por Hedge [23,24]. Este algoritmo recebe como parâmetros de entrada o número dos classificadores utilizados e o parâmetro de aprendizagem $\beta \in [0,1]$. A cada classificador são atribuídos um peso (w_i) e uma *probabilidade de distribuição normalizada* (p_i), que significa a probabilidade de acerto. Os pesos são actualizados em tempo real após cada decisão, isto é, atribuição de uma classe. Na primeira iteração do algoritmo, não existem dados estatísticos suficientes para escolher um classificador, por isso, é seleccionado aleatoriamente um classificador inserido no conjunto de classificadores. Se este errar a classificação, a sua *perda individual* toma o valor 1 e caso contrário toma o valor 0. Após cada decisão, o peso do classificador é actualizado. Se o classificador realizar uma decisão correcta o seu peso mantém-se, se realizar uma decisão incorrecta o seu peso diminui. Na iteração seguinte, o primeiro passo a realizar é recalcular a probabilidade da distribuição normalizada, onde os classificadores correctos aumentam a sua probabilidade de distribuição. Ao fim de bastantes iterações, a probabilidade de distribuição revela o classificador mais exacto (figura 3.3).

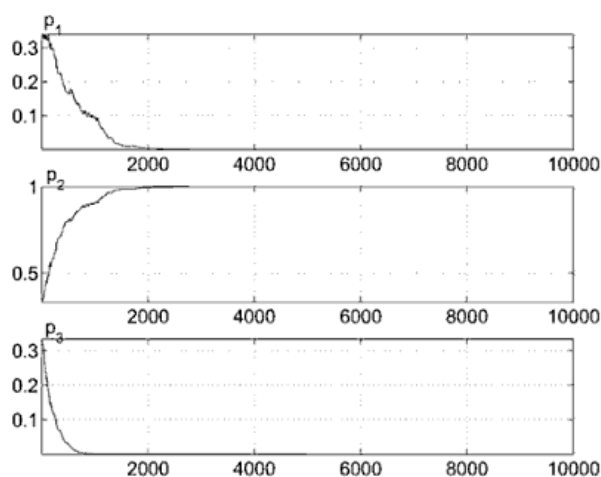


Figura 3.3: Alterações na distribuição sobre um conjunto de 3 classificadores utilizando o algoritmo de Hedge [24].

De forma a melhorar a visualização global do algoritmo, a figura 3.4 mostra todos os passos do algoritmo durante a sua execução [24]:

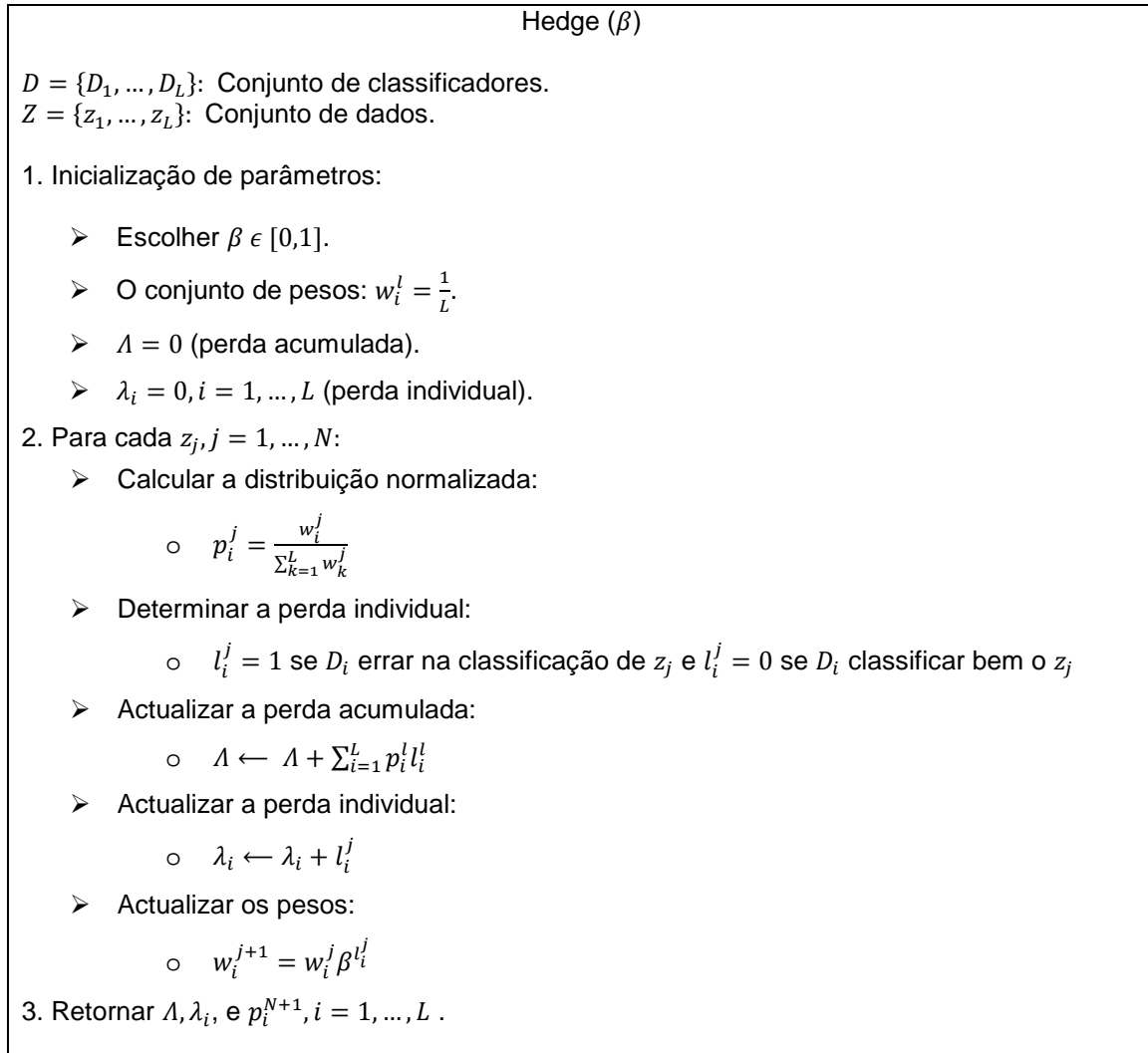


Figura 3.4: Algoritmo de Hedge [24].

O algoritmo de Hedge actualiza a distribuição (p_i^j) de forma a minimizar a perda acumulada. O valor de β é calculado através das seguintes equações:

$$\beta = g(\tilde{\lambda}/\ln L) \tag{3.15}$$

$$g(\zeta) = \frac{1}{1 + \sqrt{\frac{2}{\zeta}}} \tag{3.16}$$

onde $\tilde{\lambda}$ é a estimativa do limite máximo da perda do melhor classificador do conjunto de classificadores.

Assim a perda total do algoritmo tem como limite máximo a seguinte expressão:

$$\Lambda \leq \min_{i=1} \lambda_i + \sqrt{2\tilde{\lambda} \ln L} + \ln L \quad (3.17)$$

Através da expressão 3.17, é evidente que o limite superior da perda acumulada do algoritmo de Hedge é maior do que o limite superior da perda do melhor classificador presente no conjunto de classificadores [24]. O método de Boosting teve como base de desenvolvimento ideias aplicadas neste algoritmo.

3.2.2.2. Adaboost

O Adaboost (*Adaptative Boosting*) é uma variante do Boosting, introduzido em 1995 por Freund e Schapire e, tal como o nome indica, ajusta-se adaptativamente aos erros retornados pelo classificador fraco, ao longo das iterações. A variante Adaboost pode ser utilizada na selecção das características mais discriminativas e no treino do classificador, com a intenção de otimizar a classificação binária.

O conjunto de treino (x_i, y_i) é constituído pelos valores dos padrões (x_n) e pelas suas respectivas classificações *a priori* (y_n) . O Adaboost executa repetidamente o classificador fraco, seleccionando e combinando as melhores *hipóteses* de separação entre as duas classes e as respectivas características mais discriminativas, de forma a gerar um único e eficaz classificador forte. O classificador fraco tem o objectivo de determinar as hipóteses de separação, ao longo do vector de características, que retornam o menor erro na distinção binária entre as duas classes. A característica associada à melhor hipótese de separação é seleccionada. O classificador fraco $(h_j(x))$ depende do valor da característica do padrão (f_j) , do *threshold* (θ_j) e da paridade que indica a direcção da equação (p_j) :

$$h_j(x) = \begin{cases} 1 & \text{se } p_j f_j(x) < p_j \theta_j \\ 0 & \text{caso contrário} \end{cases} \quad (3.18)$$

Na inicialização do classificador fraco, todos os padrões têm o mesmo peso (w_i) . Após a finalização da primeira iteração, selecciona-se a melhor hipótese de separação e, os pesos dos padrões são actualizados consoante a classificação resultante da primeira iteração. Os padrões correctamente classificados diminuem o seu peso e os padrões erradamente classificados mantêm o peso da iteração anterior. Sendo assim, na próxima iteração os padrões com maior peso possuem maior influência na localização da hipótese de separação, pois o objectivo do classificador fraco em cada iteração é obter o inferior erro. No final das iterações o classificador forte utiliza o valor dos pesos para calcular alguns parâmetros (β_t, α_t) decisivos na obtenção da classificação final. Na figura 3.5 apresentam-se todos os passos na execução do Adaboost [19].

Adaboost

- Dados os padrões de entrada com as suas respectivas classificações *a priori* $(x_i, y_i) \dots (x_n, y_n)$, onde $y_i = 0, 1$ para a classe *A* e *B*, respectivamente.
- Inicialização dos pesos $w_{t,i} = 1/m$, onde m é o número total de padrões de entrada.
- Para $t = 1, \dots, T$
 1. Normalização dos pesos,
 - $w_{t,i} = \frac{w_{t,i}}{\sum_{i=1}^n w_{t,i}}$;
 2. Treino do classificador fraco na selecção da melhor hipótese (h_j), por cada característica j . O erro é avaliado através do w_i ,
 - $E_j = \sum_i w_i |h_j(x_i) - y_i|$;
 3. Escolher a hipótese h_j com menor erro E_j ;
 4. Actualização dos pesos,
 - $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$, onde $\beta_t = \frac{e_t}{1-e_t}$;
 - Onde $e_i = 0$ se o padrão x_i foi correctamente classificado, $e_i = 1$ caso contrário.

- Classificador forte,

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{caso contrário} \end{cases} \quad \alpha_t = \log \frac{1}{\beta_t}$$

Figura 3.5: Todos os passos na execução da variante Adaboost [19].

3.2.2.2.1. Análise do erro de treino

Uma propriedade importante do Adaboost consiste na capacidade de reduzir rapidamente o erro de treino. Qualquer que seja a hipótese de separação determinada pelo classificador fraco, entre classes equivalentes e binárias, o erro de classificação é sempre igual ou inferior a $\frac{1}{2}$. Sendo assim, é possível escrever o erro como, $\epsilon_t = \frac{1}{2} - \gamma_t$, onde γ_t é o valor de melhoramento em relação ao erro obtido pela pior hipótese possível ($\epsilon_t = \frac{1}{2}$).

O erro de treino obtido pelo classificador forte através do Adaboost tem como limite superior a seguinte expressão [23]:

$$\epsilon \leq \prod_{t=1}^T 2\sqrt{\epsilon_t(1-\epsilon_t)} = \prod_{t=1}^T \sqrt{(1-4\gamma_t^2)} \quad (3.19)$$

Se o erro de treino de cada hipótese do classificador fraco for ligeiramente melhor do que o erro máximo, é possível concluir que o erro de treino decresce exponencialmente:

$$\epsilon \leq \prod_{t=1}^T \sqrt{1 - 4\gamma_t^2} \leq \exp\left(-2 \sum_{t=1}^T \gamma_t^2\right) \quad (3.20)$$

Na figura 3.6 é possível visualizar a evolução do erro de treino no Adaboost em função do número de iterações. Assim, se o conjunto de teste for similar ao conjunto de treino podem prever-se óptimos resultados.

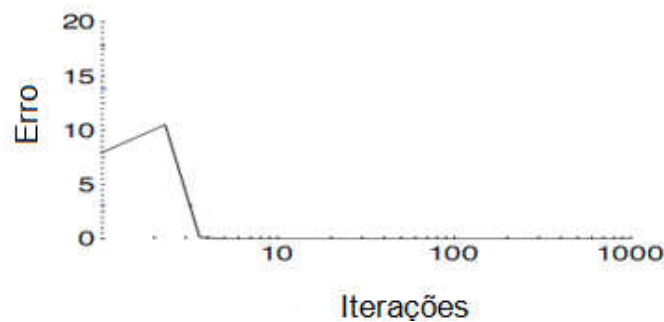


Figura 3.6: O rápido decrescimento do erro de treino em função do número de iterações [25].

3.2.2.2. Exemplo do Adaboost

Na figura 3.7 representam-se duas classes distintas, a Azul e a Vermelha, com 125 padrões por cada classe. Através da execução cíclica do classificador fraco são obtidas as melhores hipóteses de separação, h_t . As hipóteses de separação diferem de iteração para iteração, devido às constantes actualizações dos pesos ao longo da execução do Adaboost.

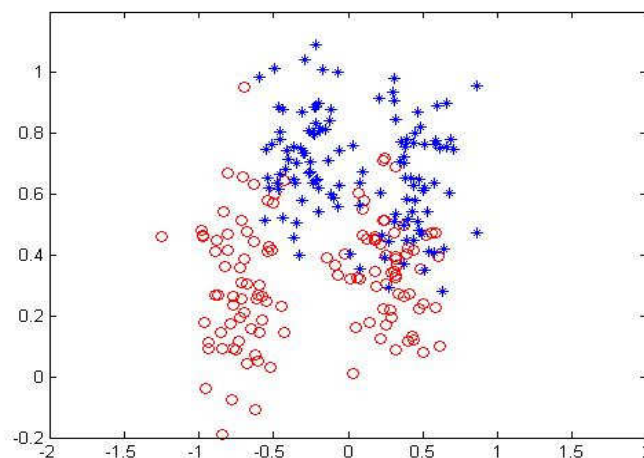


Figura 3.7: Disposição dos padrões a classificar.

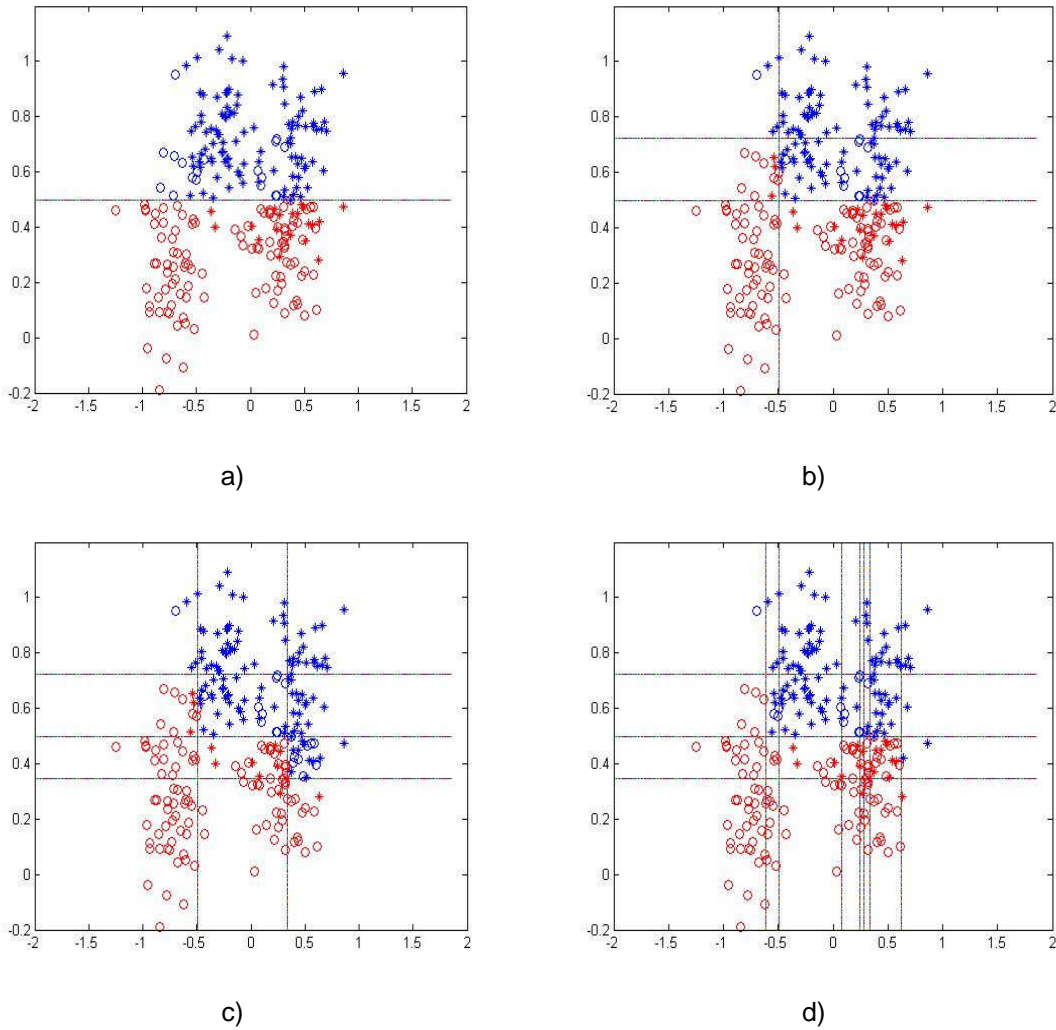


Figura 3.8: Execução do Adaboost. a),b),c) e d) classificador final após a realização de 1,3,5 e 10 iterações, respectivamente.

Inicialmente, na figura 3.7, os padrões possuem o mesmo peso, $w_{1i} = 1/250$. O classificador fraco obtém o erro mínimo na separação entre as duas classes com a hipótese horizontal (figura 3.8 a)). A primeira hipótese classifica erradamente 37 padrões, originando um $e_1 = 0.148$, $\beta_1 = 0.174$ e $\alpha_1 = 0.760$.

Após a primeira iteração, os padrões bem classificados diminuem o valor dos seus pesos, enquanto que os padrões incorrectamente classificados mantêm os seus pesos. Nas iterações seguintes, o classificador fraco pretende corrigir os padrões erradamente classificados para obter o erro mínimo. Este exemplo possui alguns padrões de classes distintas muito próximos, não possibilitando a visualização do decaimento exponencial do erro do conjunto de treino. Mesmo assim, na tabela seguinte verifica-se o decaimento suave do erro com o aumento das iterações.

Nº de Iterações	Erro obtido %
1	14.8
3	12.8
5	12.4
10	12.0
20	8.8
30	8.4
50	7.2
100	3.6

Tabela 3.1: Decaimento do erro do conjunto de treino

3.2.2.2.3. Generalização do Adaboost

Na realização de experiências utilizando o Adaboost, verificou-se a tendência de não existir sobreaprendizagem, após a execução de centenas ou milhares de iterações. Mais especificamente, utilizando as árvores de decisão como classificador fraco no reconhecimento de caracteres [26], concluiu-se que a sobreaprendizagem não esteve presente.

3.2.2.3. Classificação de multi-classes

Em muitos casos há necessidade de atribuir uma classe entre várias classes. Para obter classificações correctas nessas circunstâncias, foram geradas extensões ao Adaboost. Nas próximas secções explicam-se sucintamente as variantes Adaboost.M1 e Adaboost.M2.

3.2.2.3.1. Adaboost.M1

A primeira extensão do Adaboost, de nome Adaboost.M1, é extremamente simples [23]. No Adaboost.M1, a classificação *a priori* dos padrões de entrada pertence a um conjunto de k possíveis classes, $Y = \{0,1,2,3,\dots,k\}$. Uma das premissas subjacente à classificação binária consiste no erro mínimo na separação entre duas classes, conter o valor inferior ou igual a $\frac{1}{2}$. A partir desta constatação é possível provar (3.20), que o erro de treino decresce exponencialmente. Na classificação multi-classe esta premissa não é uma realidade. Durante a execução do Adaboost.M1, se o erro $\epsilon_t > \frac{1}{2}$, a iteração corrente é abortada e o classificador fraco retorna a calcular outra hipótese de separação. Na eventualidade do classificador fraco não conseguir determinar nenhuma hipótese de separação com um erro $\epsilon_t \leq \frac{1}{2}$, o Adaboost.M1 termina.

Por outro lado, o classificador forte atribui a classe mais frequente ao padrão desconhecido x , através da seguinte expressão:

$$h(x) = \arg \max_{y \in Y} \sum_{t=1}^T \alpha_t (h_t(x) = y) \quad (3.21)$$

3.2.2.3.2. Adaboost.M2

A segunda extensão ao Adaboost, promove a comunicação entre o algoritmo de Boosting e o classificador fraco, com o fundamento de tornar o classificador fraco mais flexível aos erros resultantes das hipóteses de separação das classes distintas [23]. Tal como no Adaboost.M1, existe mais de duas classes diferentes, por isso, os rótulos dos padrões de entrada correspondem a um intervalo finito, $Y = \{1, 2, \dots, k\}$.

A grande inovação do Adaboost.M2 relativamente ao Adaboost.M1 consiste, em procurar um melhor rótulo para os padrões incorrectamente classificados pelo classificador fraco. Sendo assim, cada padrão erradamente classificado é posteriormente testado com os restantes rótulos, de forma a verificar qual o melhor rótulo para esse determinado padrão. Cada rótulo possui um determinado peso, $q(i, y)$, de forma a distinguir a importância dos rótulos na atribuição ao padrão incorrectamente classificado.

Para verificar a precisão das hipóteses do classificador fraco é definido o “pseudo-perda”:

$$\epsilon_t = \frac{1}{2} \sum_{i=1}^N D_t(i) \left(1 - h_t(x_i, y_i) + \sum_{y \neq y_i} q_t(i, y) h_t(x_i, y) \right) \quad (3.22)$$

onde $h_t(x_i, y_i) = [0, 1]$, e $D_t(i)$ é a normalização dos pesos dos padrões mal classificados.

Neste algoritmo, o classificador fraco não se foca apenas na obtenção das melhores hipóteses, mas também na correcção dos rótulos incorrectos, que são extremamente difíceis de eliminar. Com estas novas características o classificador fraco do Adaboost.M2 consegue atingir hipóteses com erros $\epsilon_t \leq 1/2$.

Em suma, o método de Boosting é um método capaz de melhorar a precisão de qualquer classificador, especialmente dos classificadores fracos. Relativamente à variante Adaboost, não há necessidade de conhecer o funcionamento do classificador fraco, nem é necessário atribuir valores de ajuste como parâmetros de entrada. Possui uma excelente propriedade inexistente noutros classificadores, a capacidade de identificar os padrões incorrectamente rotulados, que geram ambiguidades na estrutura do classificador. Por vezes, existem alguns padrões com pesos elevados e, o classificador fraco constantemente ao longo das iterações não consegue gerar hipóteses capazes de classificar correctamente esses padrões, levando ao

aumento dos seus pesos. Esses padrões podem, eventualmente, possuir rótulos incorrectos. Contudo a característica que realça o Boosting e as suas variantes é a capacidade de resistir à sobreaprendizagem.

Por outro lado, o Adaboost tem desvantagens, como a dependência de uma boa escolha do classificador fraco. Se o classificador fraco possuir um desempenho medíocre ao longo das iterações, o classificador forte tende a ser pior. Também, o desempenho do Adaboost degrada-se na presença de características ambíguas, devido às taxas altas de erro do classificador fraco. De referir, a complexidade computacional elevada na selecção das melhores características sobre um conjunto numeroso de características ao longo das iterações.

4. Resultados Experimentais

Neste capítulo, apresentam-se os resultados obtidos pelos métodos dos k -vizinhos mais próximos e de Boosting na detecção da doença de Alzheimer em imagens PET. Os métodos foram aplicados num conjunto de 195 imagens PET, divididas igualmente em três classes (AD, Normal e MCI). O objectivo deste capítulo é analisar o desempenho dos métodos dos k -vizinhos mais próximos e de Boosting no diagnóstico automático da doença de Alzheimer. Por outro lado, no final avalia-se as potencialidades destes métodos no diagnóstico automático da doença de Alzheimer. Antes de apresentar os resultados obtidos, explica-se o pré-processamento e a origem das imagens PET.

4.1. Base de Dados ADNI

O aumento da incidência da doença de Alzheimer e a dificuldade em descobrir a cura para a demência mais frequente, propiciou um grande investimento na tentativa de diagnosticar a doença de Alzheimer num estado precoce, de forma a evitar sua progressão na deterioração do tecido nervoso.

O site do *Alzheimer's Disease Neuroimaging Initiative* (ADNI) [27] contém uma base de dados de vários géneros de imagens, onde sobressaem as imagens PET, MRI e FMRI, oriundas de bastantes pacientes nas classes Alzheimer, MCI e Normal. O ADNI recrutou aproximadamente 800 pacientes para realizar vários exames durante dois anos, dos quais 200 pacientes pertenciam à classe Normal, 400 pacientes à classe MCI e, os restantes 200 pacientes pertenciam à classe AD. Bianualmente, as imagens dos aproximadamente 800 pacientes foram actualizadas, de forma a identificar a evolução da deterioração do tecido nervoso. As imagens da base de dados ADNI são de acesso restritivo aos investigadores previamente admitidos. O ADNI tem a função de ajudar os investigadores no descobrimento de novos tratamentos para anular a evolução da doença de Alzheimer num estado precoce. No mesmo sentido, o ADNI tem como objectivo principal a identificação das regiões do cérebro e as alterações cognitivas associadas à doença de Alzheimer num estado precoce.

Na base de dados ADNI foram retirados vários volumes cerebrais de imagens PET, que não reportavam qualquer anomalia resultante de erradas execuções no exame PET. O número máximo de volumes de imagens PET, correspondente a diferentes pacientes na classe AD foi de 65. Nas outras classes existiam maiores quantidades, contudo também foram retirados da base de dados ADNI a mesma quantidade de volumes de imagens PET. Não foram utilizadas as imagens PET intervaladas no tempo do mesmo paciente, de forma a não introduzir redundância.

Cada volume da imagem PET tem a dimensão 128 por 128 e possui 60 cortes axiais ao longo do cérebro (“slices”). No total foram retirados 195 volumes de imagens PET pré-processadas, divididos igualmente pelas três classes e com as seguintes características apresentadas na tabela 4.1.

	AD	Normal	MCI
Nº Imagens	65	65	65
Sexo	40M/25F	43M/22F	44M/21F
Idade Média	76.994	76.660	77.318
Desvio Padrão	6.846	4.688	7.635
MMSE	21.154 (8min/27max)	29.323(24min/30max)	26 (20min/30max)
Peso Médio	76.598	75.985	75.566
Desvio Padrão	13.997	12.924	13.138

Tabela 4.1: Apresenta-se das características das imagens PET

De referir que os valores do teste cognitivo MMSE podem oferecer um diagnóstico errado do paciente. Neste conjunto de imagens PET, há pacientes na classe AD que possuem resultados no teste cognitivo MMSE na escala do Normal. Também é de salientar, os valores médios elevados do MMSE na classe AD e MCI, isto é, apesar de estarem afectados pela doença de Alzheimer, apenas apresentam sintomas ligeiros na falha das funções cognitivas.

4.2. Pré-processamento das Imagens PET

As 195 imagens PET oriundas da base de dados ADNI são sujeitas a processos complexos de processamento, para tornar as imagens provenientes de diferentes scanners em imagens uniformes e com o mesmo formato. Após o pré-processamento é possível utilizar as intensidades dos voxels no treino do classificador. Por vezes no momento do exame PET, o paciente encontra-se incorrectamente posicionado no aparelho e, juntamente com as anatomias e morfologias distintas do cérebro entre os pacientes, origina que a mesma localização de um voxel em duas imagens PET diferentes, corresponda a divergentes posições anatómicas no cérebro. Para corrigir esse problema é necessário realizar a *normalização espacial*. Por outro lado, durante um exame PET são extraídas várias perspectivas cerebrais, orientadas pelo sistema de coordenadas *Talairach*. Após a certificação da correcta sobreposição do conjunto de voxels que constituem as imagens PET na anatomia do cérebro, é necessário escalar as intensidades dos voxels através da *normalização da intensidade*. Distintas quantidades de FDG presentes no momento da aquisição das imagens, a sensibilidade do scanner e a diferente evolução da patologia entre os pacientes, causam desnivelamentos nas intensidades dos voxels. Finalmente realiza-se a suavização das imagens com o *Full Width Half Maximum (FWHM)*. Na suavização, a imagem PET passa por um filtro que elimina o ruído e consequentemente melhora a qualidade da imagem PET [10].

Após o pré-processamento das imagens PET é possível utilizar e comparar directamente as intensidades dos voxels entre as imagens PET. Na imagem 4.1 representam-se as diferenças entre uma imagem PET original e uma imagem PET pré-processada, do mesmo paciente sobre a mesma localização axial.

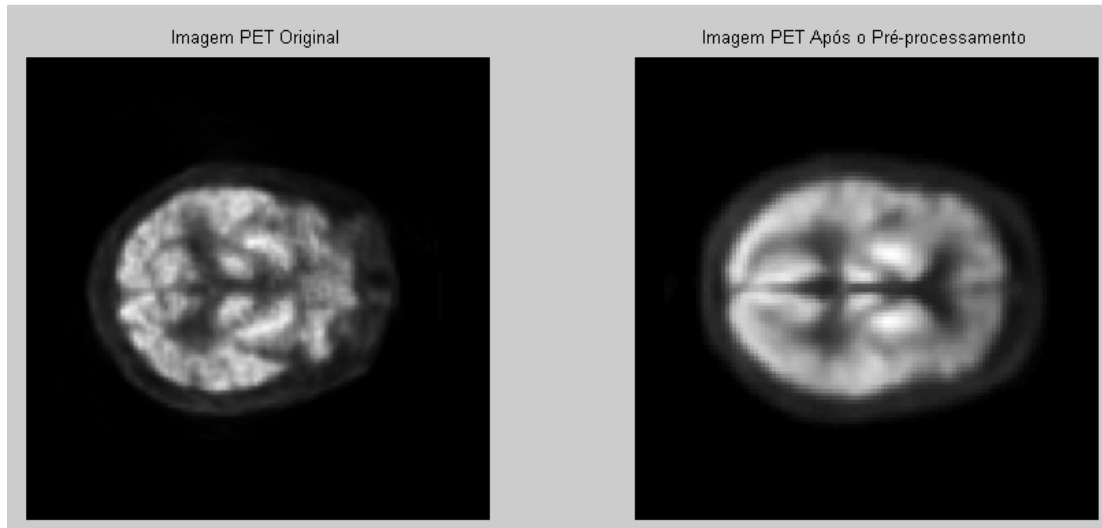


Figura 4.1: Diferença entre a imagem original e a imagem Pré-processada.

4.3. Utilização de Máscaras

O volume da imagem PET possui informações irrelevantes e de nula preponderância no reconhecimento da doença de Alzheimer. As informações irrelevantes apenas incrementam a ambiguidade e decrementam a precisão na classificação. Das 60 slices axiais ao longo cérebro, existem slices que são extra crânio e que não retêm nenhuma informação na identificação da doença de Alzheimer, como é possível visualizar na figura 4.2. Por estas razões, não foram utilizadas todas as slices disponíveis, mas sim, apenas as slices 15 à 45.

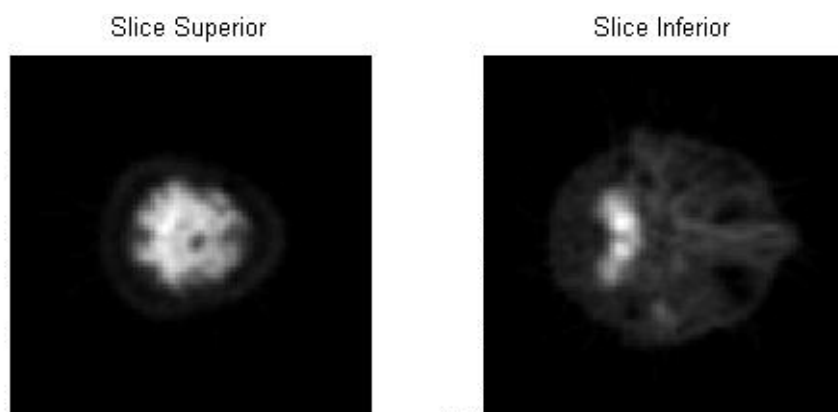


Figura 4.2: Slices irrelevantes no reconhecimento da doença do Alzheimer.

Para além das slices não utilizadas, e como já foi referido anteriormente, existem determinadas regiões do cérebro não propícias ao desenvolvimento da doença de Alzheimer. As máscaras têm a função de eliminar essas regiões, e assim, inserir apenas as intensidades dos voxels emergentes das regiões preponderantes na identificação da doença de Alzheimer no vector de características. Foram utilizadas dois tipos de máscaras, designadas por máscara M1 e máscara M2, com funções diferentes. A máscara M1 engloba praticamente a totalidade do cérebro, eliminando o lobo frontal posterior e as extremidades circundantes do cérebro. A máscara M2 elimina o lobo frontal, as extremidades circundantes e o “corpo caloso”.

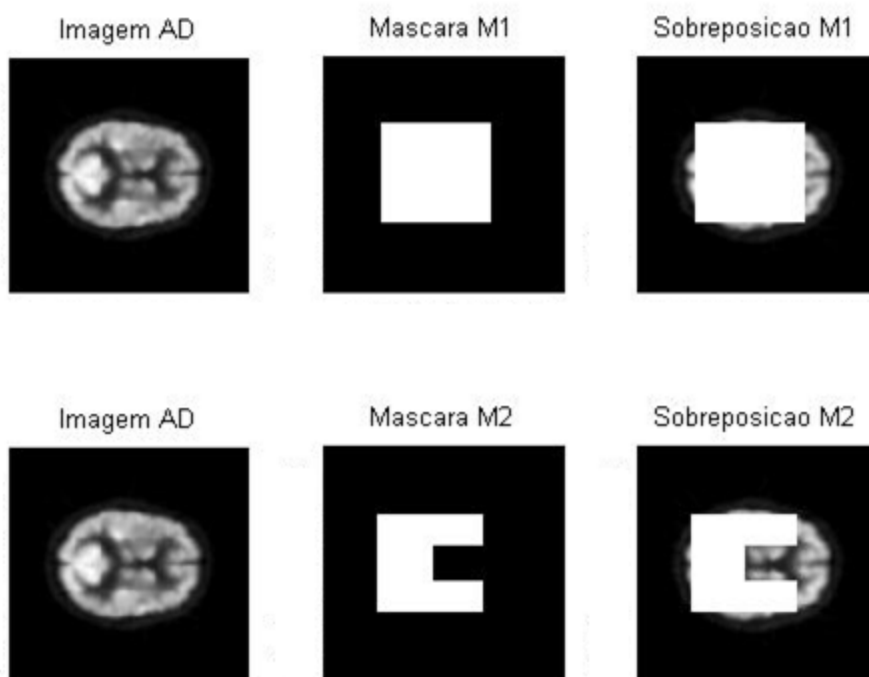


Figura 4.3: Localização das máscaras M1 e M2.

4.4. Resultados do Método dos k -Vizinhos Mais Próximos

O método dos k -vizinhos mais próximos foi o primeiro método utilizado nesta dissertação, para a identificação da doença de Alzheimer. É um método simples, que atribui a classe maioritária dos k vizinhos mais próximos à imagem PET a classificar.

4.4.1. Características Fundamentais

A imagem PET é um conjunto de voxels que contêm as intensidades da actividade cerebral. As regiões propícias ao desenvolvimento do Alzheimer têm menor actividade cerebral, assim, este facto, é a premissa para o cálculo das características mais discriminativas. Para cada imagem PET, realizou-se o somatório da distância entre cada voxel de uma imagem PET e o correspondente voxel das restantes imagens PET. Sendo assim, a

distância de uma imagem PET para outra imagem PET quantifica-se por único valor. Esta operação realizou-se entre as classes AD e Normal, AD e MCI, e finalmente entre as classes MCI e Normal. Foram aplicadas duas métricas, norma L_1 e norma L_2 (equações 4.1 e 4.2), na realização do somatório.

$$L_1 = \|I - I'\|_1 = \sum_{k,l,m} |I(k,l,m) - I'(k,l,m)| \quad (4.1)$$

$$L_2 = \|I - I'\|_2 = \sqrt{\sum_{k,l,m} (I(k,l,m) - I'(k,l,m))^2} \quad (4.2)$$

Nas regiões de maior distância, a norma L_2 obteve maiores diferenças entre as imagens de diferentes classes. Na figura 4.4 apresentam-se as matrizes das distâncias entre as classes AD e Normal com as normas L_2 e L_1 .

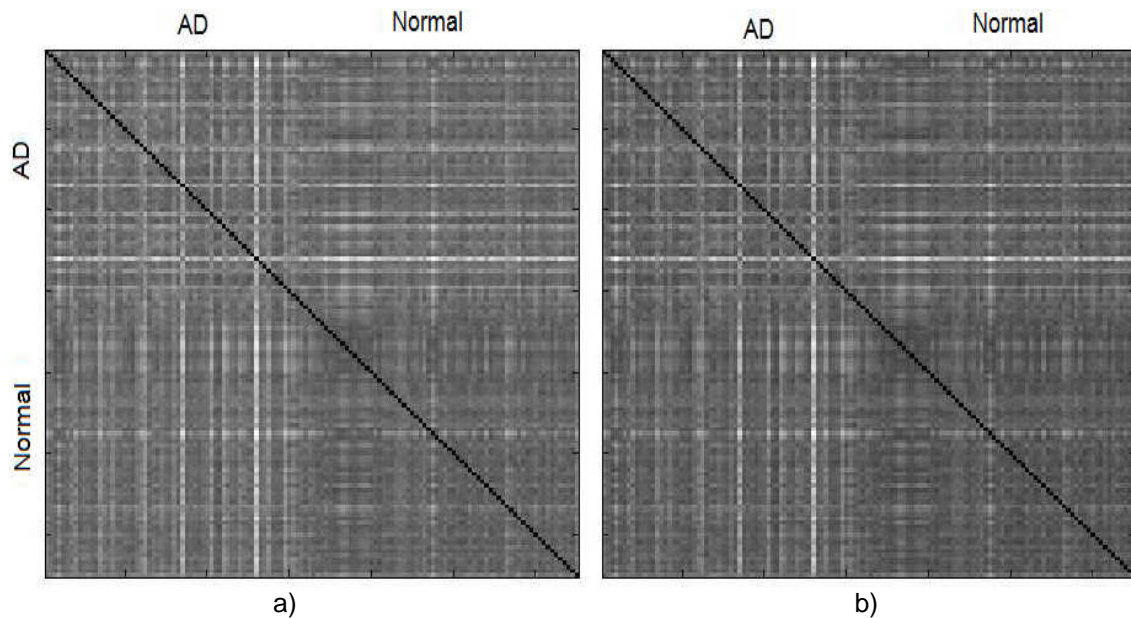


Figura 4.4: Matriz das distâncias entre as imagens das classes AD e Normal sobre a máscara M1. a) norma L_2 ; b) norma L_1 .

As matrizes das distâncias têm a dimensão de 130 por 130, em que as primeiras 65 posições pertencem à classe AD e as restantes posições competem à classe Normal. A diagonal a preto na matriz das distâncias representa a distância entre duas imagens iguais, cuja intensidade é nula. Por outro lado, a distância entre a imagem I_1 e I_2 é igual à distância entre I_2 e I_1 , logo a matriz das distâncias é uma matriz simétrica, tal como se representa na figura 4.4. Teoricamente, as imagens PET da mesma classe são bastante semelhantes, prevendo assim, valores mais baixos no retorno das distâncias entre as imagens da mesma classe do que nas distâncias entre as imagens de classes adjacentes. Na matriz das distâncias entre as classes AD e Normal, evidenciou-se uma zona menos luminosa no quadrante

Normal/Normal. Isto significa que as imagens Normais são mais homogêneas do que as imagens AD. O contraste escasso na luminosidade entre os quadrantes AD/AD e AD/Normal vai provocar ambiguidade na classificação das imagens PET, utilizando o método dos k -vizinhos mais próximos. Nas matrizes das distâncias entre as classes AD e MCI (figura 4.5) e, MCI e Normal (figura 4.6), as diferenças luminosas entre quadrantes foram menos perceptíveis. Com a inserção da classe MCI, uma classe intermédia entre as classes AD e Normal, motivou a diminuição da discrepância entre as imagens PET.

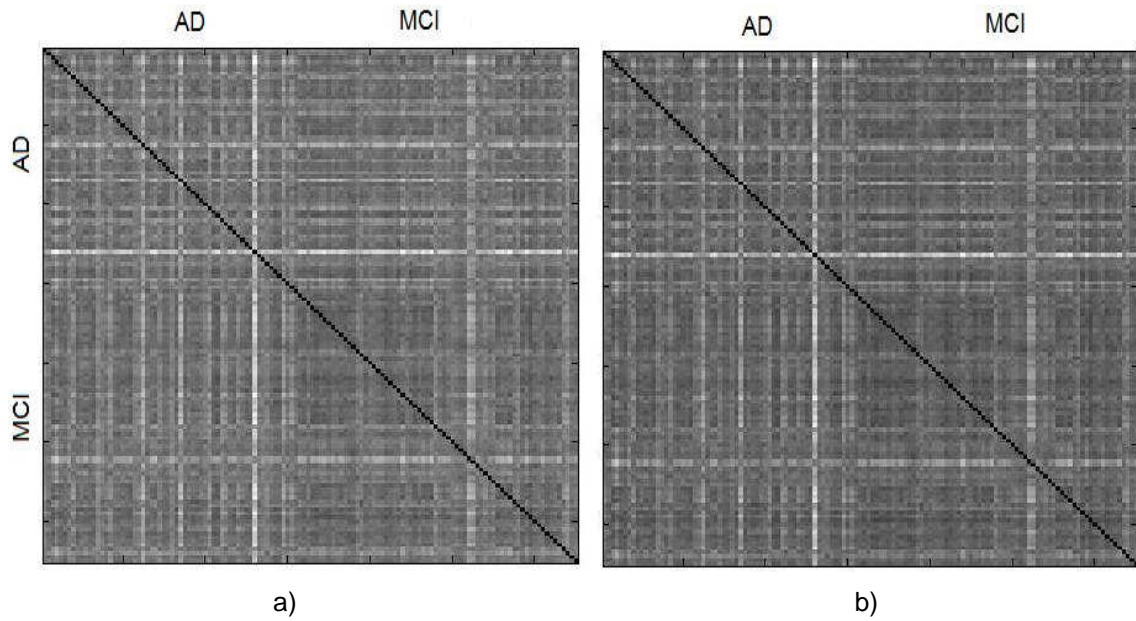


Figura 4.5: Matriz das distâncias entre as imagens das classes AD e MCI sobre a máscara M2. a) norma L_2 ; b) norma L_1 .

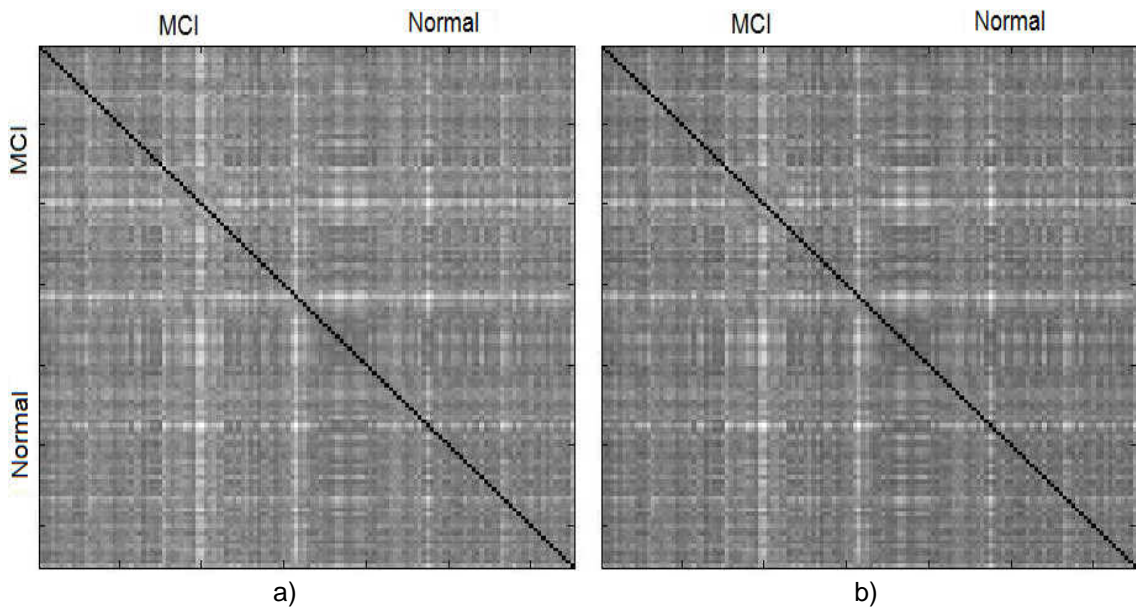


Figura 4.6: Matriz das distâncias entre as imagens das classes MCI e Normal sobre a máscara M1. a) norma L_2 ; b) norma L_1 .

4.4.2. Classificação

A imagem PET a classificar depende dos 129 valores dispostos nas colunas da matriz das distâncias. Os primeiros 65 valores pertencem à classe c e os restantes valores à classe c' . Se a imagem PET a classificar pertencer à classe c , é de esperar que os primeiros 65 valores sejam menores do que os restantes valores. Assim, após a escolha dos k vizinhos mais próximos, determinaram-se as k menores posições relativamente às 129 posições. Se a classe c foi maioritária entre as k menores posições, a imagem PET obteve a classificação correcta, se a classe c' foi maioritária entre as k menores posições, a imagem PET obteve a classificação incorrecta. Quando a imagem PET a classificar pertenceu à classe c' determinaram-se os k valores mais elevados.

4.4.3. Avaliação do Desempenho do Método dos k -vizinhos Mais Próximos

Utilizou-se a técnica da validação cruzada para realizar a avaliação do método dos k -vizinhos mais próximos. As 130 imagens PET, pertencentes a duas classes, foram divididas aleatoriamente em 5 subconjuntos de 26 imagens PET. Após a rotação do subconjunto de teste por todos os subconjuntos, calculou-se a média da Sensibilidade, Especificidade e Exactidão do método dos k -vizinhos mais próximos na distinção binária entre as classes AD e Normal, AD e MCI, e MCI e Normal.

4.4.4. Resultados e Análise

As imagens PET a classificar dependem exclusivamente das classes das k imagens mais próximas. Devido a esse facto, experimentou-se vários valores ímpares de vizinhos mais próximos, contudo evitou-se a utilização de valores elevados e muito baixos. Nos valores elevados de k , há a possibilidade de apanhar excesso de vizinhos pertencentes à outra classe, e nos valores baixos de k podem restringir a classificação, tornando-a pouco fiável. Na tabela 4.2 é possível visualizar os resultados do método dos k -vizinhos mais próximos para os melhores valores de k , sobre as intensidades dos voxels do cérebro, máscara M1 e máscara M2. Também é possível visualizar as diferenças nos resultados utilizando a norma L_1 ou a norma L_2 .

Classe		AD/Normal			AD / MCI			MCI/ Normal		
Máscara/ Métrica	Valor de K	Sens.	Esp.	Exact.	Sens.	Esp.	Exact.	Sens.	Esp.	Exact.
Cérebro L_1	5	58.46	90.77	74.62	55.38	67.69	61.54	44.62	75.38	60.00
	9	50.77	92.31	71.55	56.92	67.69	62.41	44.62	72.31	58.47
	15	55.38	90.77	73.08	64.62	60.00	62.31	35.38	76.96	56.15
Cérebro L_2	5	61.54	89.23	75.39	55.38	61.54	58.46	43.08	75.38	59.23
	9	53.85	92.31	73.08	58.46	61.54	60.00	35.38	87.69	61.54
	15	50.77	89.23	70.00	56.92	53.85	55.34	36.92	84.62	60.77
Máscara M1 L_1	5	66.15	92.31	79.23	44.62	63.07	53.85	52.31	81.54	66.93
	9	66.15	95.38	80.77	56.92	60.00	58.46	46.15	83.08	64.62
	15	67.69	95.38	81.54	53.85	58.46	56.16	40.00	89.23	64.62
Máscara M1 L_2	5	70.77	95.38	83.08	50.77	60.00	55.39	44.61	84.62	64.62
	9	70.77	96.92	83.85	53.85	66.15	60.00	52.31	80.00	66.16
	15	64.62	95.32	80.00	55.28	63.08	59.23	44.62	92.31	68.74
Máscara M2 L_1	5	64.62	93.85	79.24	52.31	66.15	59.23	40.00	83.08	61.54
	9	63.08	98.46	80.77	60.00	70.77	65.39	33.64	84.23	59.04
	15	66.15	98.46	82.31	55.38	72.31	63.85	38.46	92.31	65.39
Máscara M2 L_2	5	66.15	90.77	78.46	53.85	66.15	60.00	44.62	80.00	62.31
	9	64.62	96.92	80.77	50.77	67.69	59.23	30.77	90.77	60.77
	15	66.15	100	83.08	56.92	73.85	65.39	36.92	95.38	66.15

Tabela 4.2: Apresenta-se os resultados percentuais do método dos k -vizinhos mais próximos.

Numa apreciação global, os resultados da distinção binária sobre as máscaras atingiram melhores desempenhos do que a utilização de todos os voxels do volume do cérebro. As principais razões foram as diferenças morfológicas nas extremidades circulares do cérebro e de algumas slices irrelevantes no reconhecimento da doença de Alzheimer, que não foram incluídas nas máscaras. Também é possível visualizar que a norma L_2 obteve melhores resultados relativamente à norma L_1 , pois a norma L_2 acentua as diferenças entre as imagens de classes distintas.

Na distinção entre as classes AD e Normal atingiu-se uma exactidão de 83.85%, sobre a máscara M1 e utilizando a norma L_2 . A especificidade foi bastante elevada, tal como era de esperar, já que na matriz das distâncias existe uma perfeita diferença de luminosidade no quadrante Normal/Normal. Na distinção entre as classes AD e MCI a exactidão decresceu para 65.39%, sobre a máscara M2 e usando a norma L_2 . Alcançaram-se valores baixos de especificidade e sensibilidade, assim, com a inclusão da classe MCI a dificuldade de distinguir correctamente a doença de Alzheimer em estados de evolução distintos foi elevada. O aumento da semelhança entre as imagens das duas classes levou ao decréscimo da exactidão. Na distinção entre as classes MCI e Normal, a exactidão foi um pouco superior do que a distinção entre as classes AD e MCI, atingindo o valor de 68.74% sobre a máscara M1 e a norma L_2 . Este valor deveu-se essencialmente a baixa sensibilidade, isto é, muitas imagens MCI foram classificadas como imagens Normais. Como já foi referido, a classe MCI é um estado de transição entre o estado Normal e o estado Alzheimer, o que dificulta a discriminação entre o estado MCI e os estados adjacentes.

Os erros na discriminação entre as classes surgiram porque as imagens PET inseridas na mesma classe não são globalmente semelhantes. Esta ambiguidade nas intensidades dos voxels proporcionou falsidade na classificação das imagens PET. Na figura 4.7 ilustra-se a dissemelhança entre imagens da mesma classe.

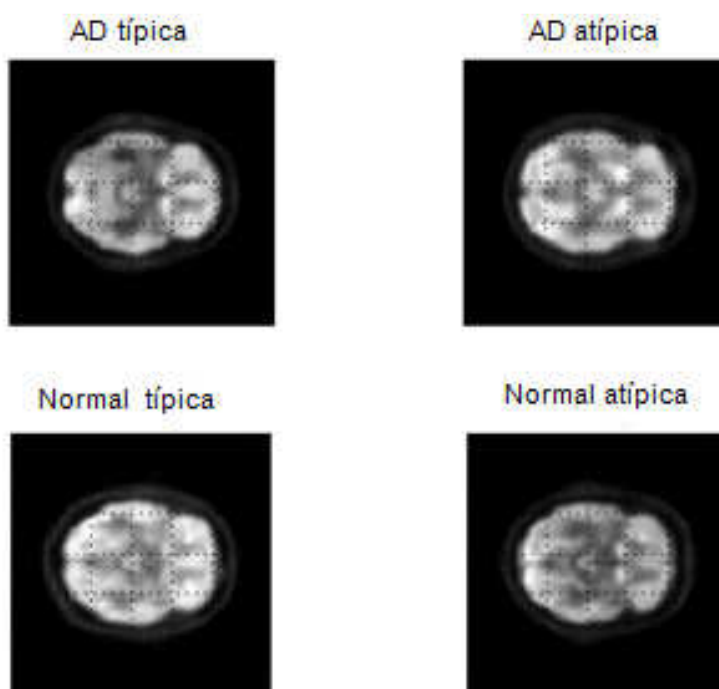


Figura 4.7: Dissemelhança entre imagens da mesma classe.

4.4.5. Localização das Principais Regiões Indicadas pelo *k*-Vizinhos Mais Próximos

Há determinadas regiões do cérebro mais propícias ao desenvolvimento da doença de Alzheimer. Através das distâncias entre as imagens PET de classes diferentes, é possível visualizar as regiões do cérebro mais distintas e conseqüentemente as regiões mais prováveis na identificação da doença de Alzheimer. A fim de obter o parecer de um especialista na análise de imagens PET, o Dr. Imagiologista António Bastos Leite analisou as regiões de maiores distâncias ao longo das slices 5 à 55. De salientar, duas slices que possuem maior visibilidade no reconhecimento da doença de Alzheimer.

Na representação da slice 20, o Dr. António Bastos Leite assinalou as regiões mais salientes no reconhecimento da doença de Alzheimer, e comentou “que as zonas parietais são bastante boas, contrariamente as diferenças centrais que se devem ao alargamento ventricular”. Na representação da slice 35, o Dr. António Bastos Leite tornou a inserir as regiões mais preponderantes na distinção da doença de Alzheimer e, comentou “apenas as ovais são boas, e que o resto é extra crânio, sendo pior quanto mais superior for a slice”.

Nas duas figuras que se seguem representam-se as regiões do cérebro mais distintas na discriminação entre as classes AD e Normal sobre as slices 20 e 35.

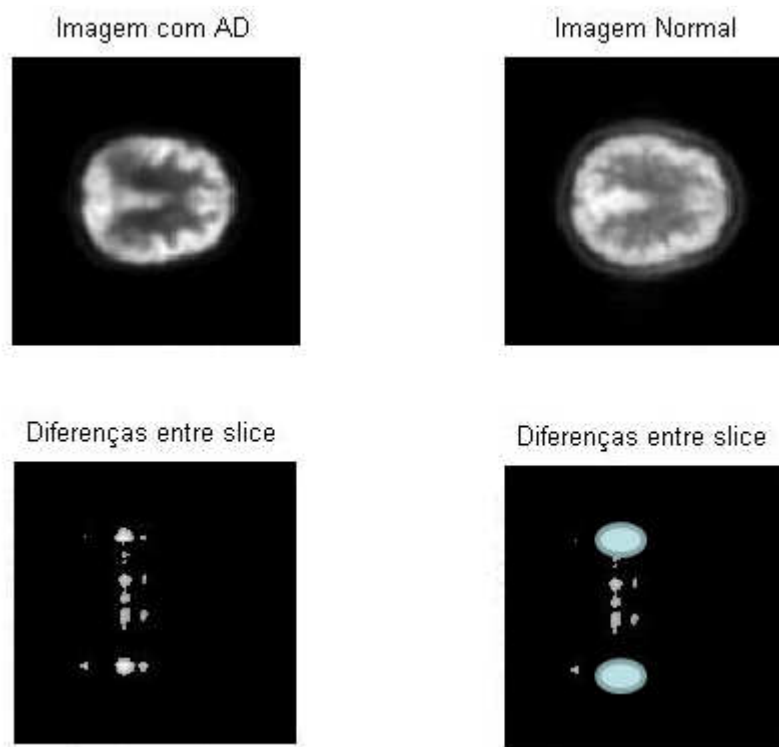


Figura 4.8: Localização das regiões mais distintas na slice 20.

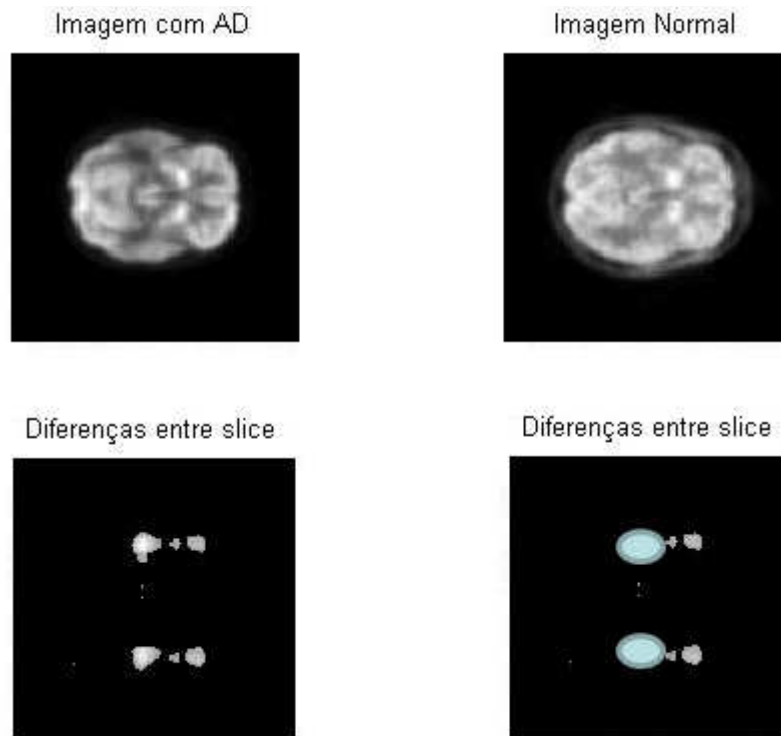


Figura 4.9: Localização das regiões mais distintas na slice 35.

4.5. Resultados do Adaboost

Após a utilização do método dos k -vizinhos mais próximos, simples e intuitivo, optou-se por um método mais complexo e com capacidade de se adaptar ao problema da “maldição da dimensionalidade”. O método de Boosting é um método capaz de melhorar a exactidão de qualquer classificador, especialmente dos classificadores fracos. Para classificar as imagens PET escolheu-se o Adaboost, variante do método de Boosting, que possui a vantagem de ser resistente à sobreaprendizagem.

4.5.1. Características Fundamentais

Cada imagem PET possui milhares de voxels. Contudo é necessário reduzir o número de voxels a inserir no vector de características na obtenção uma classificação precisa. Utilizou-se o Adaboost, na selecção das características fundamentais.

Utilizaram-se duas técnicas para a extracção de características das imagens PET:

- Os valores das intensidades dos voxels: Contêm o valor isolado da intensidade cerebral. Sobre a máscara M1 e a máscara M2 o vector de características possui 79856 e 73160 valores de intensidades, respectivamente.

- Features de Haar e a sua média: Contêm o valor médio e Haar de um volume do cérebro. Sobre a máscara M1 e a máscara M2 o vector de características possui 124296 e 109472 valores de intensidades, respectivamente. De referir que as features de Haar percorrem todas as posições das máscaras, e apenas posteriormente são seleccionados os melhores volumes na distinção binárias entre as classes. Na figura 4.10 representa-se a forma e as dimensões tridimensionais das features de Haar utilizadas.

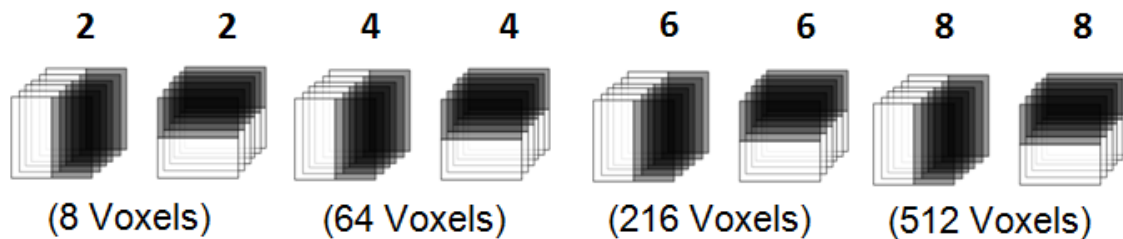


Figura 4.10: A forma e as dimensões tridimensionais das features de Haar utilizadas na extracção de características [28].

O Adaboost seleccionou as melhores características ao longo da execução cíclica do classificador fraco, de forma a obter um preciso classificador forte. Assim, o Adaboost executou ciclicamente os seguintes 4 passos (na secção 3.2.2.2 explica-se pormenorizadamente o algoritmo):

- 1) Normalização dos pesos de cada imagem. Inicialmente os pesos são equivalentes;
- 2) O classificador fraco determinou a melhor hipótese de separação para todas as posições do vector de características;
- 3) Escolheu-se a hipótese com menor erro na distinção binária das classes e seleccionou-se a característica correspondente;
- 4) Actualização dos pesos. A imagem incorrectamente classificada mantém o peso e a imagem correctamente classificada diminui o peso.

O ciclo Adaboost foi executado por 300 vezes, em todas as situações (distinção binária entre classes, sobre a máscara M1 e M2, utilizando como características as features de Haar e as intensidades dos voxels). A execução do Adaboost na selecção de características foi de elevada carga computacional e temporal, já que o Adaboost percorreu centenas de milhares de posições do vector de características, de forma a determinar a melhor característica para a actual distribuição dos pesos. Cada iteração do ciclo demorou aproximadamente 9 minutos, logo cada ciclo de 300 iterações demorou cerca de 45 horas. O Adaboost foi implementado em Matlab e correu em vários computadores *duo core* para tornar a simulação mais rápida. O desempenho do classificador forte melhora com o aumento de iterações por ciclo, pois a variante Adaboost é resistente à sobreaprendizagem, contudo escolheu-se 300 iterações como um número bastante elevado e computacionalmente possível de se realizar. Assim, o vector de

características sofreu uma enorme redução para apenas 300 posições, capazes de identificar e classificar com precisão a doença de Alzheimer.

4.5.2. Classificação

O vector de características, derivado da execução cíclica do classificador fraco, contém 300 características mais discriminativas e os seus respectivos valores de confiabilidade (alfa α_t). Para cada imagem PET de teste, o classificador forte percorreu as 300 características seleccionadas e, no final, atribuiu uma classificação à imagem PET, segundo a seguinte equação:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{t=300} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{t=300} \alpha_t \\ 0 & \text{caso contrário} \end{cases} \quad (4.3)$$

Se o valor do rótulo da imagem PET de teste, y_i , coincidiu com o retorno do classificador forte, a imagem PET respectiva obteve a classificação correcta, se os valores não coincidiram, a imagem PET obteve a classificação incorrecta.

4.5.3. Avaliação do Desempenho do Adaboost

A validação cruzada foi a técnica utilizada na avaliação do Adaboost. As 130 imagens PET, pertencentes a duas classes, foram divididas aleatoriamente em 5 subconjuntos de 26 imagens PET, igualmente partilhadas pelas respectivas classes. Após a rotação do subconjunto de teste foi possível obter os resultados do Adaboost na distinção binária entre as 3 classes. De salientar, a sobrecarga computacional e temporal na execução da validação cruzada na avaliação do Adaboost. A realização do treino sobre o conjunto de 104 imagens PET demorou cerca de 45 horas. Como se realizaram 5 rotações no conjunto de treino, o treino completo correspondeu a vários dias. Por outro lado, ainda se realizaram 3 distinções binárias sobre as características das features de Haar e das intensidades dos voxels, o que fez passar os dias para meses.

4.5.4. Resultados e Análise

Após a execução do classificador forte sobre as 300 características mais discriminativas na distinção entre as classes, atingiu-se os resultados do Adaboost. Na tabela 4.3 representam-se os resultados do Adaboost utilizando as características das intensidades dos voxels e das features de Haar, sobre as máscaras M1 e M2.

Características/ Máscara	AD/Normal			AD / MCI			MCI/ Normal		
	Sens.	Esp.	Exact.	Sens.	Esp.	Exact.	Sens.	Esp.	Exact.
Voxels/ M1	83.08	87.69	85.39	64.54	58.46	61.54	63.08	75.38	69.23
Voxels/ M2	80.00	92.31	86.15	69.23	63.08	66.15	58.46	73.85	66.15
Haar/M1	87.69	87.69	87.69	61.54	52.31	56.96	64.62	76.92	70.77
Haar/M2	86.15	92.31	89.23	53.84	58.46	56.15	69.23	63.08	66.08

Tabela 4.3: Apresenta-se os resultados percentuais do Adaboost.

Antes de analisar os resultados do Adaboost, de referir que, tal como já se relatou na secção 3.2.2.2.1, se verificou que o erro do treino decresce exponencialmente. Com apenas 10 a 20 iterações, na maioria das situações, o erro de treino é nulo.

Durante a execução do classificador fraco determinam-se as características, que possuem menor erro na discriminação entre classes distintas. Por isso, as imagens do conjunto de treino que possuem maior peso, resultante de classificações erradas, têm maior preponderância na escolha das características nas próximas iterações, de forma atingir o menor erro na discriminação. Assim, a selecção das características flutua entre as dimensões elevadas do vector de características, para manter os pesos das imagens reduzidos, o que leva ao rápido decréscimo do erro de treino.

O facto das imagens AD e Normais pertencerem a dois estados bastantes distintos, resultou na selecção de características mais discriminativas capazes de identificar correctamente os dois estados. Assim, a distinção entre as classes AD e Normal atingiu uma exactidão elevada de 89.23%. De referir, tal como aconteceu no método dos *k*-vizinhos mais próximos, a especificidade foi superior à sensibilidade, logo, as imagens PET Normais são mais fáceis de reconhecer.

Na distinção entre as classes AD e MCI, uma distinção entre estados mais próximos, a exactidão decresceu para o valor de 66.15%. Em sincronia com a exactidão, a sensibilidade e a especificidade também diminuíram os seus valores percentuais. Principalmente com a utilização das features de Haar existe uma dispersão elevada entre os valores dos volumes Haar, ao longo das diferentes imagens PET nas mesmas localizações. Devido à grande dispersão no conjunto de treino e, a conseqüente mistura dos valores pertencentes às duas classes, provoca que a melhor hipótese de separação atinja uma percentagem elevada de imagens classificadas incorrectamente. Na distinção entre as classes MCI e Normal, à

semelhança do que aconteceu na distinção entre as classes AD e MCI, o valor percentual da exactidão diminuiu para 70.77%.

A falta de homogeneidade das imagens PET dentro da mesma classe e a falta de heterogeneidade das imagens PET entre as classes distintas dificulta a realização de uma classificação precisa. Uma forma de diminuir este problema é reduzir a dimensão da máscara, de forma a extrair apenas as regiões mais propícias ao desenvolvimento da doença de Alzheimer e consequentemente obter as regiões mais distintas na discriminação entre as classes, como por exemplo, utilizar apenas a intensidade dos voxels do hipocampo ou lobo temporal.

Por outro lado, a dimensão elevada do vector de características, antes da utilização do Adaboost na selecção de características, pode provocar ambiguidade na sua execução. O classificador fraco, ao longo da execução do Adaboost, determina as hipóteses de separação com menor erro na distinção binária entre as classes. Contudo nesta procura exaustiva do menor erro ao longo das centenas de milhares de posições do vector de características, o classificador fraco pode encontrar o menor erro em regiões não propícias ao desenvolvimento da doença de Alzheimer e ambíguas, que apenas diminuem o erro de separação porque classificam bem as imagens com pesos elevados. Assim, para atenuar este acontecimento, reduziu-se a dimensão do vector de características, através de uma selecção inicial. Esta selecção consiste em determinar os volumes Haar ou os voxels que possuem características mais discriminativas entre as classes (características com as maiores diferenças entre a média dos seus 52 valores pertencentes a cada classe). Desta forma, o Adaboost executou a selecção de características sobre uma dimensão reduzida do vector de características.

Optou-se por reduzir em 1/5 a dimensão original do vector de características, sobre as condições que obtiveram *a priori* os melhores desempenhos na discriminação entre as classes. Percorreu-se todas as posições do vector de características e seleccionou-se 1/5 das características mais discriminativas entre as classes distintas. Na tabela 4.4 é possível visualizar os resultados desta variante.

Distinção binária	Características / Máscara	Redução 1/5	Sens.	Esp.	Exact.
AD/Normal	Haar/ M2	109472/20000	86.15	84.62	85.38
AD / MCI	Voxels/ M2	73160/15000	67.69	67.69	67.69
MCI/ Normal	Haar/ M1	124296/24000	63.08	76.92	70.00

Tabela 4.4: Apresenta-se os resultados percentuais do Adaboost com redução inicial do vector de características.

Globalmente, os resultados com a redução do vector de características foram semelhantes. Apenas de salientar, que na distinção entre as classes AD e MCI atingiu-se um ligeiro aumento da exactidão. Por outro lado, a redução do vector de características provocou uma diminuição enorme da sobrecarga computacional.

4.5.5. Localização das Principais Regiões Indicadas pelo Adaboost

A utilização do Adaboost teve, numa primeira instância, a selecção das características mais discriminativas capazes de aumentar a precisão e generalização do classificador forte. Ao longo das 300 iterações, o Adaboost percorreu as centenas de milhares de posições no vector de características para seleccionar os melhores volumes Haar na identificação da doença de Alzheimer.

Na figura 4.12 é possível visualizar as principais 6 regiões Haar determinadas pelo Adaboost sobre as respectivas slices. Utilizou-se as features de Haar porque obtiveram melhores desempenhos do que as intensidades dos voxels. De referir, tal como era previsto, as principais 6 regiões determinadas pelo Adaboost encontram-se no lobo parietal e temporal. A região mais seleccionada encontra-se na slice 16 e sobrepõem-se na região do hipocampo (figura 4.11).

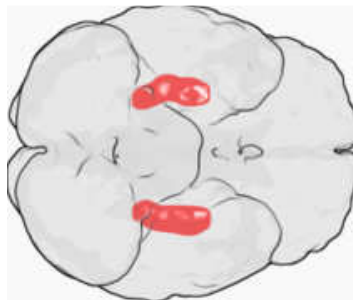


Figura 4.11: Localização do hipocampo [29].

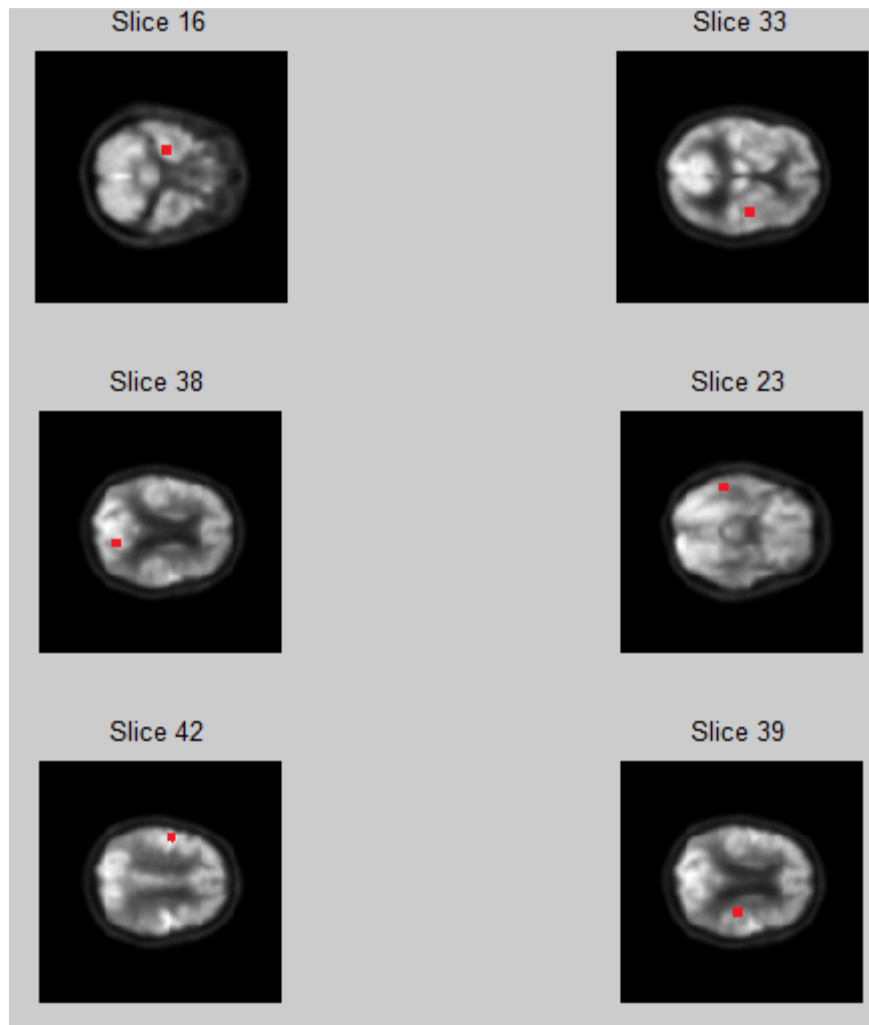


Figura 4.12: As principais 6 regiões determinadas pelo Adaboost.

4.6. Comparação dos Resultados

Em todas as distinções binárias entre as classes, o Adaboost atingiu melhores desempenhos na exactidão relativamente ao k -vizinhos mais próximos. A resistência à sobreaprendizagem e capacidade de adaptação ao problema da “maldição da dimensionalidade” foram factos preponderantes, e que prevaleceram durante a execução do Adaboost. Na tabela 4.5 mostra-se a comparação entre os melhores resultados obtidos pelo Adaboost e k -vizinhos mais próximos.

Classificadores	Adaboost			<i>k</i> -vizinhos mais próximos		
	Sens.	Esp.	Exact.	Sens.	Esp.	Exact.
AD/Normal	86.15	92.31	89.23	70.77	96.92	83.85
AD / MCI	67.69	67.69	67.69	56.92	73.85	65.39
MCI/ Normal	64.62	76.92	70.77	44.62	92.31	68.74

Tabela 4.5: Comparação dos resultados dos classificadores (%).

5. Conclusão

A doença de Alzheimer é actualmente a demência mais frequente, que deteriora progressivamente o tecido nervoso do cérebro humano. As consequências da doença de Alzheimer correspondem a perda gradual da memória e das funções cognitivas (linguagem, orientação, juízo crítico), culminando numa fase terminal, na perda total de autonomia. Passados cem anos desde a descoberta da demência, ainda não existe cura, apenas tratamentos que diminuem a progressão da doença de Alzheimer.

A dificuldade na interpretação e análise visual das imagens PET no diagnóstico da doença de Alzheimer motivaram a utilização dos métodos dos k -vizinhos mais próximos e de Boosting no diagnóstico automático da doença de Alzheimer nas imagens PET. As virtudes do método de Boosting prevaleceram e conseqüentemente permitiram atingir melhores desempenhos em todas as distinções binárias entre as classes.

A discriminação entre a classe AD e a classe Normal obteve melhores desempenhos do que as restantes distinções binárias. O facto dos estados AD e Normal não serem estados transitórios, possibilitou encontrar características mais discriminativas e assim obter uma exactidão da distinção de 89.23%.

Nesta dissertação considerou-se o estado MCI, ainda pouco estudado em sistemas automáticos, mas essencial no diagnóstico precoce da doença de Alzheimer. O MCI é um estado de transição entre o estado Normal e o estado Alzheimer, que em muitos casos corresponde à situação de Alzheimer precoce. O facto de ser um estado de transição provocou a diminuição da exactidão, atingindo 70.77% e 67.69% na distinção entre as classes MCI e Normal, e as classes AD e MCI, respectivamente.

Em suma, o método de Boosting, através da variante Adaboost, atingiu melhores desempenhos na identificação da doença de Alzheimer em imagens PET, e possui sérias potencialidades de se tornar numa ferramenta de diagnóstico automático da doença de Alzheimer.

5.1. Trabalhos Futuro

Com a realização desta dissertação e no encadeamento das conclusões descritas anteriormente, sugere-se como trabalho futuro os seguintes tópicos:

- Utilização de máscaras mais selectivas e restritas sobre as regiões propícias ao desenvolvimento da doença de Alzheimer, tendo em vista o aperfeiçoamento da distinção binária entre as classes. Por outro lado, este facto conduzirá à redução do vector de características e à diminuição da carga computacional e temporal.
- Aplicação de outro classificador fraco no Adaboost. As árvores de decisão e o SVM são dois bons candidatos.
- Comparação entre as sensibilidades das imagens PET e MRI na identificação da doença de Alzheimer no estado precoce, moderado e acentuado.

Anexo A

Neste Anexo apresentam-se as funções do córtex cerebral. O córtex cerebral é constituído pelos lobos temporal, frontal, parietal e occipital.

O **lobo temporal** é uma região no cérebro que apresenta um significativo envolvimento com a memória. O lobo temporal localiza-se abaixo do osso temporal (acima das orelhas). Há claras evidências que interligam a região do lobo temporal com o armazenamento de eventos passados. As áreas associativas do lobo temporal estão envolvidas no reconhecimento, identificação e nomeação dos objectos. Danos no lobo temporal causam deficiências, como por exemplo, o indivíduo recebe estímulos mas não os consegue identificar. Algumas destas deficiências podem ser bastantes específicas, como a incapacidade de reconhecer faces, mesmo reconhecendo a pessoa através de outras características, como a voz. O lobo temporal contém o neocórtex temporal, região potencialmente envolvida na memória a longo prazo. Nesta região também há um grupo de estruturas interconectadas entre si que exercem funções da memória para factos e eventos, entre elas está o hipocampo. O **hipocampo** é uma estrutura localizada nos lobos temporais do cérebro humano, considerada a principal zona da memória e importante componente do sistema límbico. Além disso, está relacionada com a orientação espacial. Estudos indicam que o hipocampo é fundamental para converter a memória de curto prazo em memória de longo prazo.

O **lobo frontal** está envolvido nas habilidades motoras (incluindo a fala) e nas funções cognitivas. O centro motor do cérebro (giro pré-central) localiza-se na retaguarda do lobo frontal e à frente do lobo parietal. Recebe e processa conexões da parte somatossensorial do lobo parietal e inicia as funções motoras. As funções associativas do lobo frontal estão relacionadas com o planeamento. Danos a este nível causam perturbações de personalidade.

O **lobo parietal** recebe e processa todas as entradas somatossensoriais do corpo (toque, dor). As fibras da medula espinhal distribuem-se pelo tálamo para várias partes do lobo parietal. Essas conexões formam um "mapa" da superfície do corpo no lobo parietal. Esse mapa é chamado de homúnculo. A retaguarda do lobo parietal (próxima ao lobo temporal) possui uma secção, de nome *área de Wernicke*, muito importante para compreender as informações sensoriais (visuais e auditivas) associadas à linguagem. Danos nessa área do cérebro produzem a "afasia sensorial", na qual os pacientes não conseguem entender a linguagem mas ainda são capazes de produzir sons.

O **lobo occipital** recebe e processa informação visual. As áreas associativas estão relacionadas com a interpretação do mundo visual e do transporte da experiência visual na comunicação. Danos no lobo occipital podem causar deficiências muito específicas, como por exemplo, não ser capaz de detectar movimento numa imagem.

Anexo B

Mini Mental State Examination (MMSE)

1. Orientação (1 ponto por cada resposta correcta)

Em que ano estamos? _____

Em que mês estamos? _____

Em que dia do mês estamos? _____

Em que dia da semana estamos? _____

Em que estação do ano estamos? _____

Nota:_____

Em que país estamos? _____

Em que distrito vive? _____

Em que terra vive? _____

Em que casa estamos? _____

Em que andar estamos? _____

Nota:_____

2. Retenção (contar 1 ponto por cada palavra correctamente repetida)

"Vou dizer três palavras; queria que as repetisse, mas só depois de eu as dizer todas; procure ficar a sabê-las de cor".

Pêra _____

Gato _____

Bola _____

Nota:_____

3. Atenção e Cálculo (1 ponto por cada resposta correcta. Se der uma errada mas depois continuar a subtrair bem, consideram-se as seguintes como correctas. Parar ao fim de 5 respostas)

"Agora peço-lhe que me diga quantos são 30 menos 3 e depois ao número encontrado volta a tirar 3 e repete assim até eu lhe dizer para parar".

27_ 24_ 21 _ 18_ 15_

Nota:_____

4. Evocação (1 ponto por cada resposta correcta.)

"Veja se consegue dizer as três palavras que pedi há pouco para decorar".

Pêra _____

Gato _____

Bola _____

Nota:_____

5. Linguagem (1 ponto por cada resposta correcta)

a. "Como se chama isto? Mostrar os objectos:

Relógio _____

Lápis _____

Nota:_____

b. "Repita a frase que eu vou dizer: O RATO ROEU A ROLHA"

Nota:_____

c. "Quando eu lhe der esta folha de papel, pegue nela com a mão direita, dobre-a ao meio e ponha sobre a mesa"; dar a folha segurando com as duas mãos.

Pega com a mão direita_____

Dobra ao meio _____

Coloca onde deve_____

Nota:_____

d. "Leia o que está neste cartão e faça o que lá diz". Mostrar um cartão com a frase bem legível, "FECHE OS OLHOS"; sendo analfabeto lê-se a frase.

Fechou os olhos_____

Nota:_____

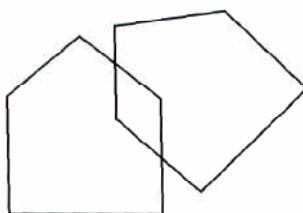
e. "Escreva uma frase inteira aqui". Deve ter sujeito e verbo e fazer sentido; os erros gramaticais não prejudicam a pontuação.

Frase:

Nota:_____

6. Habilidade Construtiva (1 ponto pela cópia correcta.)

Deve copiar um desenho. Dois pentágonos parcialmente sobrepostos; cada um deve ficar com 5 lados, dois dos quais intersectados. Não valorizar tremor ou rotação.



Cópia:

Nota:_____

TOTAL (Máximo 30 pontos):_____

Referências Bibliográficas

- [1] Luís Leite. “**Demências**”. Apontamentos de Neurologia, 2004.
- [2] http://www.encyclopedia.com.pt/articles.php?article_id=729.
- [3] <http://www.dsc.ufcg.edu.br/~pet/jornal/dezembro2007/saude.html>.
- [4] Chris Hinrichs; Vikas Singh; Lopamudra Mukherjee; Guofan Xu; Moo K. Chung; Sterling C. Johnson . “**Spatially augmented LP Boosting for AD classification with evaluations on the ADNI dataset**”. *NeuroImage*, pp. 138-149, May 2009.
- [5] Jonathan H. Morra; Zhuowen Tu; Liana G. Apostolova; Amity E. Green; Arthur W. Toga; Paul M. Thompson. “**Comparison of AdaBoost and Support Vector Machines for Detecting Alzheimer's Disease Through Automated Hippocampal Segmentation**”. *IEEE Transactions on Medical Imaging*, February 2008.
- [6] A. Savio; M.García-Sebastián; M. Graña; J. Villanúa. “**Results of an Adaboost Approach on Alzheimer's Disease Detection on MRI**” *Third International Work-Conference on the Interplay Between Natural and Artificial Computation Bioinspired Applications in Artificial and Natural Computation*, pp.114-123, 2009.
- [7] Jorge S. Marques. “**Reconhecimento de Padrões métodos estatísticos e neuronais**”. IST Press, 2005.
- [8] Daniel Silverman; Gary Small. “**PET Shows Promise in Early Detection of AD**”. *Journal of the American Medical Association*, November 2001.
- [9] P.M. Thompson; K.M Hayashi; R.A Dutton; M.C Chiang; A.D Leow; E.R Sowell; G. De Zubicaray; J.T Becker; O.L Lopez; H.J Aizenstein. “**Tracking Alzheimer’s Disease**”. *Annals of the New York Academy of Sciences Special Issue on Imaging and the Aging Brain*, December 2006.
- [10] <http://www.loni.ucla.edu/ADNI/Data/index.shtml>.
- [11] Jonathan Stoeckel; Glenn Fung. “**SVM Feature Selection for Classification of SPECT Images of Alzheimer’s Disease using Spatial Information**”. *IEEE International Conference on Data Mining*, pp. 410-417, 2005.
- [12] Tang-kai Yin, Nan-Tsing Chiu. “**Discrimination between Alzheimer's dementia and controls by automated analysis of statistical parametric maps of ^{99m}Tc-HMPAO-SPECT volumes**”. *IEEE Symposium Bioinformatics and Bioengineering*, pp.183-190, July 2004.

- [13] Tang-kai Yin; Nan-Tsing Chiu. "**Fuzzy Patterns and Classification of Functional Brain Images for the Diagnosis of Alzheimer's Disease**". *IEEE International Conference on Fuzzy Systems*, pp.161-166, May 2005.
- [14] Jonathan Stoeckel; Grégoire Malandain; Octave Migneco; Pierre Malick Koulibaly; Philippe Robert; Nicholas Ayache; Jacques Darcourt. "**Classification of SPECT Images of Normal Subjects versus Images of Alzheimer's Disease Patients**". *In 4th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, vol. 2208, pp. 666-674, 2001.
- [15] J.S. Kirppernhan; J.H. Nagel. "**Diagnostic and Modelling of Alzheimer's Disease through Neural Network Analyses of PET Studies**". *IEEE Engineering in Medicine and Biology Society*, vol.12, pp. 1449 – 1450, November 1990.
- [16] S.Duchesne; A.Caroli; C.Geroldi; C.Barillot; G.B.Frisoni; D.L.Collins. "**MRI-Based Automated Computer Classification of Probable AD Versus Normal Controls**". *IEEE Transactions on Medical Imaging*, vol. 27, pp. 509- 520, April 2008.
- [17] P.A. Freeborough; N.C.Fox. "**MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease**". *IEEE Transactions on Medical Imaging*, vol 17, pp.475-478, June 1998.
- [18] K. Fritzsche; A. Von Wangenheim; R. Dillmann; R. Unterhinninghofen. "**Automated MRI-Based Quantification of the Cerebral Atrophy Providing Diagnostic Information on Mild Cognitive Impairment and Alzheimer's Disease**". *IEEE International Symposium on Computer-Based Medical Systems*, pp.191-196, 2006.
- [19] Paul Viola; Michael Jones. "**Robust Real-time Object Detection**". *International Journal of Computer Vision*, July 2001.
- [20] Christopher M. Bishop. "**Neural Networks for Pattern Recognition**". Oxford University Press,1995.
- [21] <http://www.ppgia.pucpr.br/~soares/cursos/recpad2/public/NaoParametrico.pdf>.
- [22] Yoav Freund. "**Boosting a weak learning algorithm by majority**". *Information and Computation*, vol 121, pp.256–285, 1995.
- [23] Yoav Freund; Robert E. Schapire; "**A decision-theoretic generalization of on-line learning and an application to boosting**". *Journal of Computer and System Sciences*, vol 55, pp.119-139, August 1997.
- [24] L.I. Kuncheva. "**Combining Pattern Classifiers: Methods and Algorithms**". Wiley-Interscience, 2004.

[25] Yoav Freund; Robert E. Schapire. "**A Short Introduction to Boosting**". *Journal of Japanese Society for Artelligence*, vol 14, pp. 771-780, September 1999.

[26] Harris Drucker; Corinna Cortes. "**Boosting decision trees**". In *Advances in Neural Information Processing Systems 8*, pp.479-485, 1996.

[27] <http://www.loni.ucla.edu/ADNI>.

[28] Xinyi Cui; Yazhou Liu; Shiguang Shan; Xilin Chen; Wen Gao; Harbin. "**3D Haar-Like Features for Pedestrian Detection**". *IEEE International Conference on Multimedia and Expo*, pp.1263-1266, July 2007.

[29] <http://pt.wikipedia.org/wiki/Hipocampo>.