

UNIVERSIDAD DE GRANADA



TESIS DOCTORAL

**Un nuevo modelo de la interconectividad funcional
cerebral mediante reglas de asociación aplicado
a la detección de alteraciones neurológicas**

Febrero, 2013

Rosa María Chaves Rodríguez

Departamento de Teoría de la Señal, Telemática y Comunicaciones.
Doctorado en Sistemas Multimedia.

Editor: Editorial de la Universidad de Granada
Autor: Rosa María Chaves Rodríguez
D.L.: GR 1907-2013
ISBN: 978-84-9028-616-6

D. Javier Ramírez Pérez de Inestrosa, Doctor por la Universidad de Granada y Catedrático del Departamento Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada y

D. Juan Manuel Górriz Sáez, Doctor por la Universidad de Cádiz y la Universidad de Granada y Catedrático del Departamento de Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada,

MANIFIESTAN:

Que la presente Memoria titulada *“Un nuevo modelo de la interconectividad funcional cerebral mediante imágenes tomográficas y Reglas de Asociación para la detección de alteraciones neurológicas”*, presentada por Rosa María Chaves Rodríguez para optar al grado de Doctor Internacional por la Universidad de Granada, ha sido realizada bajo nuestra dirección.

La doctoranda Rosa M^a Chaves Rodríguez y los directores de la tesis Javier Ramírez Pérez de Inestrosa y Juan Manuel Górriz Sáez garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por la doctoranda bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones. Con esta fecha, autorizamos la presentación de la misma.

Granada, a 11 de Febrero de 2013

Fdo: Javier Ramírez Pérez de Inestrosa Fdo: Juan Manuel Górriz Sáez

Memoria presentada por Rosa María Chaves Rodríguez para optar al Grado de Doctor Internacional por la Universidad de Granada.

Fdo: Rosa María Chaves Rodríguez

Título de Doctor con Mención Internacional

Con el fin de obtener la Mención Internacional en el Título de Doctor (aprobada en Junta de Gobierno de la Universidad de Granada el 28 de enero 2011, Real Decreto 99/2011), se han cumplido, en lo que atañe a esta Tesis Doctoral y a su Defensa, los siguientes requisitos:

- Durante su etapa de formación, el doctorando ha realizado dos estancias de 3 meses fuera de España en instituciones de enseñanza superior de otros Estados cursando estudios o realizando trabajos de investigación que le han sido reconocidos por el órgano responsable del programa.
- Parte de la Tesis se ha redactado y presentado en una de las lenguas habituales para la comunicación científica en su campo de conocimiento, distinta a cualquiera de las lenguas oficiales en España. En concreto en Inglés.
- La Tesis ha sido informada por dos expertos doctores pertenecientes a alguna institución de educación superior o instituto de investigación pertenecientes a alguna institución de educación superior o instituto de investigación no española.
- Un experto perteneciente a alguna institución de educación superior o centro de investigación no española, con el título de doctor, y distinto del responsable de la estancia mencionada en el apartado a), haya formado parte del tribunal evaluador de la tesis.

*“Para que surja lo posible es preciso intentar
una y otra vez lo imposible”*

Hermann Hesse

*“Si te sientas en el camino, ponte de frente a lo que todavía
te queda por andar y de espaldas a lo que ya has andado.”*

Proverbio Chino

Agradecimientos

Esta Tesis ha supuesto un proyecto personal difícil que no hubiera sido posible sin las personas que día a día me han brindado su apoyo, comprensión y que han creído en mí.

Quiero dar las gracias a mis directores por el esfuerzo continuo de corrección y guía en este trabajo.

Agradezco a todos mis compañeros de la Universidad de Granada así como a los de fuera de nuestras fronteras, que han contribuido a que hoy pueda escribir estas líneas.

Doy las gracias a mi familia, en especial a mis padres por su apoyo incondicional y su generosidad.

Gracias a tí por hacerme creer que de alguna forma con este trabajo salvaba el mundo.

Rosa Chaves Rodríguez

Financiación

Las principales fuentes de financiación de esta Tesis Doctoral han sido las siguientes:

- **Contrato Homologado en Cofinanciación con Proyectos**
Entidad financiadora: Ministerio de Educación y Ciencia. Plan Propio Universidad de Granada
Duración: Junio 2009- Mayo 2013
- **Detección de enfermedades neurológicas mediante clasificación y separación de señales** (Proyecto MEC PET2006-0253)
Entidad financiadora: Ministerio de Educación y Ciencia
Duración: 2007-2010
- **Diagnóstico avanzado de enfermedades neurológicas mediante técnicas de reconstrucción y modelado de imágenes tomográficas cerebrales** (Proyecto de excelencia P07-TIC-02566)
Entidad financiadora: Junta de Andalucía
Duración: 2008-2012
- **Toolbox for Biomedical Signal and Image Processing based on Information Theory** (Ref.: HD2008-0029)
Entidad financiadora: Ministerio de Ciencia e Innovación
Duración: 2009-2010
- **Nuevas Técnicas de Reconstrucción, Procesado, Clasificación y Fusión de Imágenes Médicas para Diagnóstico Precoz de la Enfermedad de Alzheimer** (Proyecto TEC2008-02113/TEC)
Entidad financiadora: Ministerio de Ciencia e Innovación
Duración: 2009-2012
- **Plataforma abierta de procesamiento de imágenes para ayuda al diagnóstico de alteraciones neurológicas** (Proyecto de excelencia PAPI-ADAN, P09-TIC-4530)
Entidad financiadora: Junta de Andalucía
Duración: 2010-2014
- **Estancia de investigación para la obtención del Doctorado Europeo**
Entidad financiadora: Ministerio de Educación, Cultura y Deporte
Duración: 3 meses, julio 2010-octubre 2010.

- **Estancia de Doctorado en el marco PAP-Erasmus**

Entidad financiadora: Unión Europea

Duración: 3 meses, junio 2011-septiembre 2011.

Publicaciones

Publicaciones de primer autor en revistas indexadas por el ISI derivadas de la realización de esta Tesis Doctoral

1. **R. Chaves**, J. M. Górriz, J. Ramírez, I. A. Illán, D. Salas-Gonzalez and M. Gómez-Río, Efficient mining of association rules for the early diagnosis of Alzheimer's disease, *Phys. Med. Biol.* 56, 6047, 2011.
2. **R. Chaves**, J. Ramírez, J.M. Górriz, I.A. Illán, for the Alzheimer's Disease Neuroimaging Initiative, Functional brain image classification using association rules defined over discriminant regions, *Pattern Recognition Letters*, Volume 33, Issue 12, 1 September 2012, Pages 1666-1672.
3. **R. Chaves**, J. Ramírez, J.M. Górriz, C.G. Puntonet, for the Alzheimer's Disease Neuroimaging Initiative, Association rule-based feature selection method for Alzheimer's disease diagnosis, *Expert Systems with Applications*, Volume 39, Issue 14, 15 October 2012, Pages 11766-11774.
4. **R. Chaves**, J. Ramírez, J.M. Górriz, I. A. Illán, M. Gómez-Río, C. Carnero for the Alzheimer's Disease Neuroimaging Initiative, Effective Diagnosis of Alzheimer's Disease by means of Large Margin-based methodology, *BMC medical informatics and decision making*, 12:79, 2012.
5. **R. Chaves**, J. Ramírez, J.M. Górriz and for the Alzheimer's Disease Neuroimaging Initiative, Integrating Discretization and Association Rule-based Classification for Alzheimer's Disease Diagnosis, *Experts Systems with Applications*, vol. 40, num. 5, 2013.
6. **R. Chaves**, J. Ramírez, J.M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting, *Neuroscience Letters*, Volume 461, Issue 3, 15 September 2009, Pages 293-297.

Publicaciones en actas de congresos derivadas de la realización de esta Tesis Doctoral

1. **R.M. Chaves Rodríguez**, J. Ramírez Pérez De Inestrosa, J.M. Górriz Sáez, D. Salas González, M.M. López Pérez, I. Álvarez Illan, F. Segovia Román and A. Olivares Vicente, "Effective diagnosis of Alzheimer's disease by means of distance metric learning", "Sixth international conference on hybrid artificial intelligence systems (hais'11)", 148-155, 2011.

2. **R.M. Chaves Rodríguez**, J. Ramírez Pérez De Inestrosa, J.M. Górriz Sáez, D. Salas González and M.M. López Pérez, "Distance metric learning as feature reduction technique for the Alzheimer's disease diagnosis", "Iwinac 2011 proceedings", 68-76, 2011.
3. **R.M. Chaves Rodríguez**, J. Ramírez Pérez De Inestrosa, J.M. Górriz Sáez, I. Alvarez Illan, F. Segovia Román and A. Olivares Vicente, "Effective diagnosis of alzheimer's disease by means of distance metric learning and random forest", "Iwinac 2011 proceedings", 59-67, 2011.
4. **R.M. Chaves Rodríguez**, J. Ramírez Pérez De Inestrosa, J.M. Górriz Sáez, M.M. López Pérez, D. Salas González, I. Álvarez Illán, F. Segovia Román and P. Padilla De La Torre, "Effective diagnosis of alzheimer's disease by means of association rules", "Hybrid artificial intelligence systems", 452-459, 2010.
5. **R.M. Chaves Rodríguez**, J. Ramírez Pérez De Inestrosa, J.M. Górriz Sáez, M.M. López Pérez, I. Álvarez Illán, D. Salas González and F. Segovia Román, "Spect image classification based on nmse feature correlation weighting and svm", "Proceedings of the ieee nuclear science symposium and medical imaging conference", 2715-2719, 2009.
6. **R. Chaves**, J. Ramírez, J. M. Górriz, I.A. Illán, D. Salas-Gonzalez for the Alzheimer's Disease Neuroimaging Initiative, FDG and PIB Biomarker PET analysis for the Alzheimer's disease detection using Association Rules, "Proceedings of the ieee nuclear science symposium and medical imaging conference", 2576-2579, 2012.
7. **R. Chaves**, J. Ramírez, J.M. Górriz, F. Segovia for the Alzheimer's Disease Neuroimaging Initiative, Effective voxel selection method over FDG-PET database based on Association Rules, 16th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 2210 - 2219, 2012.

Capítulo de libro

1. **R. Chaves**, D. Salas-González, J. Ramírez, J. M. Górriz, M. López, I. Álvarez, F. Segovia, Functional Brain Image Preprocessing for Computer Aided Diagnosis Systems, Recent Advances in Biomedical Signal Processing, 95-106, ISBN 978-1-60805-218-9, Editorial Bentham Science Publishers, 2011.

Abreviaturas

$A\beta$:	Amiloyd Beta (Beta Amiloide)
AChEIs:	acetylcholinesterase inhibitors (inhibidores de acetilcolinestara)
Acc:	accuracy (exactitud)
AD:	Alzheimer's Disease (Enfermedad de Alzheimer)
AE:	Activation Estimation (Estimación de la Activación)
APP:	Amiloid Precursor Proteine (Proteína Precursora del Amiloide)
APOE:	Apolipoprotein E (Apolipoproteína E)
ADAS:	Alzheimer's Disease Assessment Scale (Escala de Evaluación para la Enfermedad de Alzheimer)
ADCS-ADL:	Alzheimer's Disease Cooperative Study-Activities of Daily Living (Estudio de las actividades de la vida diaria por parte de la cooperativa AD)
ADNI:	Alzheimer's Disease Neuroimaging Initiative (Iniciativa de Neuroimagen de la Enfermedad de Alzheimer)
APP:	Amiloid Protein Precursor (Proteína Precursora de Amiloide)
AUC:	Area Under Curve (Área bajo la curva)
BEHAVE-AD:	Behavioral Pathology in Alzheimer's Disease Rating Scale (Patología de comportamiento en la escala de índice de la AD)
BHE:	Hematoencephalic Barrier (Barrera Hematoencefálica)
BFS:	Breadth-first search (Búsqueda primero en amplitud)
CAD:	Computer-Aided Diagnosis (Diagnóstico Asistido por Ordenador)
CIBIC:	Clinician's Interview-Based Impression of Change (Entrevista del médico basada en la impresión de cambio)
CBA:	Classification Based on Association (Clasificación basada en asociación)
CDR:	Clinical Dementia Rating (Clasificación Clínica de las Demencias)
CMAR:	Classification based on Multiple Association Rules (Clasificación Basa en reglas de asociación múltiples)

CPAR:	Classification based on Predictive Association Rules (Clasificación basada en Reglas de Asociación Predictivas)
CSF:	Cerebrospinal Fluid (Líquido Cefalorraquídeo)
CT:	Computed Tomography (Tomografía Computarizada)
DAD:	Disability Assessment for Dementia (Escala de Evaluación de Discapacidad para la demencia)
DLB:	Dementia with Lewy Bodies (Demencia con cuerpos de Lewy)
DFS:	Depth first search (búsqueda primero en profundidad)
DSS:	Decision Support System (Sistema de soporte de decisión)
EBC:	Energy Based Metric (Métrica Basada en Energía)
FBP:	Filtered Backprojection (Retroproyección Filtrada)
FDG:	Fludeoxyglucose (Fluorodesoxiglucosa)
FE:	Feature Extraction (Extracción de características)
fMRI:	Functional Magnetic Resonance Imaging (Imágenes Funcionales de Resonancia Magnética)
FS:	Feature Selection (Selección de características)
GMM:	Gaussian Mixture Model (Modelo de Mezcla de Gaussianas)
HMPAO:	HexaMethylPropyleneAmine Oxime (HexaMetiloPropilenoAmina Oxima)
ICA:	Independent Component Analysis (Análisis de Componentes Independientes)
KNN:	k Nearest Neighbor (k Vecinos más Cercanos)
LMNN:	Large Margin Nearest Neighbor (Vecinos más cercanos de amplio margen)

LOO:	Leave One Out (Dejar uno fuera)
MAP:	Maximum a Priori (Máximo apriori)
MCI:	Mild Cognitive Impairment (Deterioro Cognitivo Leve)
minsup:	Minimum Support (Soporte Mínimo)
minconf:	Minimum Confidence (Confianza Mínima)
MMAC:	Multi-class, multi-label associative classification (clasificación asociativa multi-clase, multi-etiqueta)
MMSE:	Mini-Mental State Examination (Examen del Estado Mental Mínimo)
MNI:	Montreal Neurological Imaging (Imagen Neurológica de Montreal)
NAA:	N-acetylaspartate (N Acetil Aspartato)
NM:	Nuclear Medicine (Medicina Nuclear)
NMSE:	Normalized Minimum Squared Error (Error cuadrático mínimo normalizado)
MRI:	Magnetic Resonance Imaging (Imágenes de Resonancia Magnética)
MTL:	Medial Temporal Lobe (Lóbulo Temporal Medio)
NIA:	National Institute on Aging (El Instituto Nacional de la Edad)
NINCDS-ADRDA:	National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (Instituto Nacional de los desórdenes comunicativos y neurológicos e impacto de la Enfermedad de Alzheimer y desórdenes relacionados)
NFT:	neurofibrillary tangles (neurofibrillas)
NPI:	Neuropsychiatric Inventory (Inventario Neuropsiquiátrico)
PCA:	Principal Component Analysis (Análisis de Componentes Principales)
PDS:	Progressive Deterioration Scale (la escala de deterioro progresivo)
PET:	Positron Emission Tomography (Tomografía por Emisión de Positrones)

PIB:	Pittsburgh-B (Pittsburgh-B)
PLS:	Partial Least Squares (Mínimos Cuadrados Parciales)
PHF:	Paired Helical Filaments (Filamento helicoidal pareado)
PSN1:	Presenilin 1 (presenil 1)
PSN2:	Presenilin 2 (presenil 2)
rCBF:	regional Cerebral Blood Flow (Flujo sanguíneo cerebral regional)
ROI:	Region Of Interest (Región de Interés)
ROC:	Receiver Operating Characteristic (Característica Operativa del Receptor)
Sen:	Sensitivity (Sensibilidad)
SIB:	Severe Impairment Battery (Batería de Desajuste Severo)
Spe:	Specificity (Especificidad)
SPECT:	Single Photon Emission Computed Tomography (Tomografía Computarizada por Emisión de un Solo Fotón)
SPM:	Statistical Parametric Mapping (Mapas Estadísticos Paramétricos)
SVM:	Support Vector Machine (Maquinas de Vectores Soporte)
TID:	Transaction Identifier (identificador de Transacción)
VAF:	Voxels-As-Features (Vóxeles Como Características)

Resumen

Este trabajo está motivado por el estudio de técnicas de reconocimiento de patrones y clasificadores para la detección precoz de alteraciones neurológicas (en particular de la Enfermedad de Alzheimer, AD del inglés “Alzheimer’s Disease”) utilizando bases de datos de imagen médica funcional del tipo: tomografía computarizada por emisión de un sólo fotón (SPECT, del inglés “Single Photon Emission Computed Tomography”) y tomografía por emisión de positrones (PET, del inglés “Positron Emission Tomography”). Con este fin, proponemos sistemas de ayuda al diagnóstico basados en dos vertientes: reglas de asociación (ARs, del inglés “Association Rules”) y algoritmo de vecinos más cercanos de amplio margen (LMNN, del inglés “Large Margin Nearest Neighbors”). Por un lado, las ARs facilitan el estudio del cerebro en términos de la interconectividad de regiones de interés, una forma dinámica y novedosa que hemos investigado desde dos puntos de vista: *i*) entre diferentes tipos de atributos (continuos, discriminativos y/o discretos) con una clasificación basada en el número de reglas verificadas por cada sujeto, y *ii*) como selector de características, con posterior reducción del espacio de entrada (mediante técnicas de extracción de características) a un clasificador “kernel”-SVM. Por otro lado, el algoritmo LMNN actúa como: *i*) transformador del espacio de características para diferenciar mejor las clases bajo estudio (normal frente a Alzheimer) y proporcionar una tasa de clasificación estable para un amplio rango de características; *ii*) como técnica de extracción de características mediante una matriz rectangular (LMNN-RECT) y/o *iii*) como clasificador utilizando las distancias Euclídea, de Mahalanobis o con métodos basados en energía. Con la experimentación realizada en diferentes bases de datos se ha demostrado la efectividad y robustez de los métodos propuestos (con exactitudes superiores al 90%) comparándolos con otros existentes hasta el momento. También se ha incidido en la aportación de los diferentes biomarcadores, como la fluorodeoxiglucosa y el componente de Pittsburgh, en la detección precoz de la AD.

Contenido

I	Introducción	xxi
1	Introducción	1
1.1	Motivación	2
1.2	Estado del arte	3
1.2.1	Sistemas de ayuda al diagnóstico	4
1.3	Objetivos	6
1.4	Organización de la memoria	6
2	La Enfermedad de Alzheimer	11
2.1	Introducción	12
2.2	Histopatología de la AD	14
2.3	Genética de la AD	15
2.4	Biomarcadores para la AD	16
2.5	Diagnóstico de la AD: tests neuropsicológicos	18
2.6	Tratamiento de la AD	20
2.7	Futuro de la AD	20
II	Estado del arte	23
3	Técnicas de obtención de imágenes médicas	25
3.1	Neuroimagen estructural y funcional	26
3.2	Medicina nuclear	28
3.3	Neuroimagen funcional	30
3.3.1	SPECT cerebral de perfusión	30
3.3.2	PET	31
3.3.3	Tomografía computarizada	34
3.3.4	Cámaras para la obtención de imágenes funcionales	34
	Cámaras SPECT	34

	Cámaras PET	35
3.4	Técnicas de preprocesado de imágenes	36
3.4.1	Reconstrucción por retroproyección filtrada	37
3.4.2	MLEM	39
3.4.3	OSEM	39
3.4.4	Normalización	40
3.4.5	Normalización espacial	41
	Registro afín	42
	Regularización	43
3.4.6	Normalización en intensidad	44
4	Diagnóstico asistido por computador	47
4.1	Introducción	48
4.2	Mapas estadísticos paramétricos (SPM)	48
4.2.1	Análisis estadísticos	50
4.2.2	Obtención de los SPMs	51
4.3	Aproximaciones multivariadas	52
III	Fundamento teórico	57
5	Fundamentos teóricos	59
5.1	Métodos de selección de características	60
5.1.1	Criterio FDR	62
	Criterio discriminante lineal de Fisher	63
	Umbral FDR para la selección de vóxeles	63
5.2	Reglas de asociación	64
5.2.1	Descripción formal del problema	65
5.2.2	Algoritmos para la extracción de ARs	67
5.2.3	Algoritmo Apriori	70
5.2.4	Métodos de discretización	76
5.3	Métodos de extracción de características	78
5.3.1	PCA	78
	Kernel PCA	79
5.3.2	PLS	80
5.4	LMNN	81
5.4.1	Función de pérdidas	82
5.4.2	Transformación LMNN y LMNN-RECT	84
5.4.3	Kernel LMNN	84
5.5	Clasificadores	85

5.5.1	Clasificación basada en técnicas de extracción de ARs	85
	Clasificación basada en evaluación de reglas	86
5.5.2	SVM	87
5.5.3	LMNN	90
5.5.4	Comparativa LMNN y SVM	92
5.6	Métodos para la evaluación	92
5.6.1	Análisis ROC y AUC	94
IV	Análisis de imagen con los métodos propuestos	97
6	Análisis de imagen mediante ARs	99
6.1	Análisis de imagen médica con ARs	100
6.2	ARs entre atributos continuos	101
6.2.1	Extracción y selección de características	103
	Enmascarado en atributos continuos	103
	Estimación de la activación	106
6.2.2	Extracción de las reglas de asociación	106
6.2.3	Clasificación	108
6.3	ARs entre atributos discriminativos	109
6.3.1	Descripción del método	109
6.3.2	Enmascarado	111
6.4	ARs entre atributos discretos	111
6.4.1	Descripción del método	112
6.4.2	Enmascarado	113
6.4.3	Vóxeles seleccionados por discretización	114
6.5	ARs como selector de características	115
6.5.1	Descripción del método	115
6.5.2	Selección de vóxeles mediante ARs	117
6.5.3	Extracción de características y clasificación	117
7	Vecinos más cercanos de amplio margen	119
7.1	Introducción	120
7.2	Descripción del método	120
7.2.1	Extracción de características	123
	Enmascarado	123
	VAF	123
	Cálculo de NMSE en ROIs tridimensionales	124
	Selección de características con t-test	125
	Reducción de características	125

Truco del Kernel	127
7.2.2 Clasificación	128
V Resultados experimentales	131
8 Bases de datos utilizadas	133
8.1 SPECT	134
8.2 ADNI-PET	136
8.2.1 ¹⁸ F-FDG ADNI PET	138
8.2.2 ¹¹ C-PIB PET	140
8.2.3 Modelo de la enfermedad y progresión utilizando ADNI	142
8.3 ¹⁸ F-FDG PET Cartuja	142
9 Resultados experimentales con ARs	145
9.1 ARs entre atributos continuos	146
9.1.1 Resultados con SPECT	147
9.1.2 Resultados para PET Cartuja	155
9.1.3 Resultados para PIB y FDG	156
9.2 ARs entre atributos discriminantes	160
9.3 ARs entre atributos discretos	169
9.3.1 Experimentos con la base de datos SPECT	170
9.3.2 Experimentos con la base de datos PET	171
9.3.3 Estudio ROC	172
9.4 ARs como selector	172
9.4.1 Experimentos para SPECT	172
9.4.2 Experimentos para FDG-PET	177
9.4.3 Análisis ROC	177
10 Resultados experimentales con LMNN	183
10.1 Selección del modelo y las características	184
10.2 Evaluación mediante k pliegues	185
10.3 Experimentos con la base de datos SPECT	186
10.4 Experimentos con la base de datos PET	189
10.5 Análisis ROC	190
VI Conclusiones y discusión	197
11 Conclusiones y trabajo futuro	199
11.1 Discusión y conclusiones de este trabajo	200

11.1.1	Reglas de asociación	200
	ARs entre atributos continuos	201
	ARs entre atributos discriminativos	202
	ARs entre atributos discretos	203
	ARs como selector	204
	Conclusiones conjuntas de los métodos basados en ARs	204
11.1.2	LMNN	205
11.2	Trabajo futuro	207
VII	Summary in English	209
12	Introduction	213
12.1	Objectives and organization	214
12.2	Neurodegenerative disorders: AD	214
12.3	Functional medical images	215
12.4	Computer aided diagnosis (CAD)	216
12.5	Databases	216
12.5.1	SPECT database	216
12.5.2	FDG-PET database	218
12.5.3	PIB-PET database	220
	Preprocessing	220
12.5.4	PET Cartuja database	221
13	Association Rules	223
13.1	Introduction	224
13.2	ARs among continuous attributes	224
13.2.1	AR-mining	224
13.2.2	Input for AR-mining algorithm	225
13.2.3	AR-based classifier	226
13.2.4	Biomarker combination and ARs	226
13.3	ARs over discriminant regions	227
13.3.1	Fisher discriminant ratio	227
13.3.2	FDR-based masking	228
13.4	ARs among discrete attributes	228
13.5	AR-based feature selector	230
13.5.1	Feature Selection	230
13.5.2	ARs for feature selection	231
	Feature extraction and classification	231
13.6	Experiments	231

13.6.1	Experiments of AR among continuous attributes	231
	Experiments of biomarker-AR-based system	238
13.6.2	Experiments of AR over discriminant regions	241
13.6.3	Experiments of discretization-AR-based system	242
	Masking for discretization procedure	244
	Voxels selected by discretization	244
13.6.4	Experiments of FS-AR-based system	247
	Experiments on SPECT database	247
	Experiments on PET database	250
	ROC analysis	250
14	Large margin nearest neighbors	253
14.1	Introduction	254
14.1.1	Loss function	255
14.1.2	LMNN-RECT as feature reduction technique	255
14.1.3	Kernel LMNN	256
14.1.4	Classification:	256
	LMNN classifier	257
	SVM classifier	257
	k -fold cross-validation evaluation	258
14.2	Experiments	258
14.2.1	Experiments with SPECT	260
14.2.2	Experiments with PET	263
14.2.3	ROC analysis	263
15	Conclusions and future work	269
15.1	Conclusions	270
15.1.1	Conclusions of ARs	270
	ARs among continuous attributes	270
	ARs over discriminant regions	271
	Discretization AR-based system	271
	AR-based FS technique	272
15.1.2	LMNN-based technique	272
15.2	Future work	273

Índice de figuras

2.1	a) Costes directos e indirectos en billones de dólares en países de altos o medios ingresos, b) Número de personas con demencia en millones frente a la evolución en años desde 2010 a 2050. Fuente: [Nations, 2008]	12
2.2	Ovillos neurofibrilares y placas seniles de los pacientes con AD respecto a los normales. Figura obtenida de [Feldman, 2007].	14
2.3	Moduladores de las relaciones temporales de biomarcador relativos a una edad fija (65 años). Hipótesis APOE ϵ 4 portador (A) y no portador (B). Extraída de [Jack et al., 2010].	16
3.1	Información anatómica de la imagen. Comparación de la magnitud de los biomarcadores funcionales frente a los estructurales teniendo en cuenta el avance de la enfermedad de Alzheimer. Obtenida de [Wechalekar and Cook, 2005]	27
3.2	Etapas de los biomarcadores en la AD. a) Un individuo cognitivamente normal sin evidencia de $A\beta$ en imágenes de amiloide PET con PiB y sin evidencia de atrofia en MRI. b) Un individuo cognitivamente normal sin evidencia de atrofia neurodegenerativa en MRI pero con depósito significativo $A\beta$ en las imágenes PET. C) Un individuo con demencia y un diagnóstico clínico de la AD, un estudio positivo de imagen de amiloide PET y una atrofia neurodegenerativa en MRI. Obtenida de [Jack et al., 2010].	29
3.3	Patrones de perfusión cerebral regional con SPECT en sujeto sano (primera fila) y paciente AD (segunda fila).	32
3.4	Escáner CT mostrando un detector gantry rotativo para la adquisición de medidas en diferentes posiciones. Obtenida de [Wechalekar and Cook, 2005].	35

3.5	a) Una imagen raw SPECT para un control (de un total de 180 proyecciones obtenidas cada 2 grados), b) sinograma SPECT, c) sinograma SPECT filtrado.	37
3.6	Preprocesado de un corte coronal de la base de datos SPECT en sus fases de: a)prefiltrado, b)reconstrucción FBP, c)postfiltrado.	38
4.1	Etapas del análisis estadístico en SPM. Fuente [Friston et al., 2007]	49
4.2	Vector de contrastes [1 -1], matriz de diseño y resultados de SPM de un estudio categórico para la base de datos PET-Cartuja.	51
4.3	Método de “clustering” basado en mezcla de gaussianas. Fuente: [Segovia et al., 2010]	54
4.4	VAF para base de datos SPECT: Los valores de todos los voxels son utilizados para construir el vector de características	54
5.1	Algoritmos más utilizados para la extracción de ARs. Obtenido de [Hipp et al., 2000].	68
5.2	Ilustración de la generación de los conjuntos de items frecuentes usando el algoritmo Apriori. Fuente [Tan et al., 2005]	72
5.3	Ilustración para la optimización de la función de pérdidas. Obtenida de [Li and Lu, 2009].	82
5.4	Ejemplo de clasificación bidimensional. A la derecha, las dos clases se pueden separar en el espacio de características por un hiperplano lineal. A la izquierda, la separación entre clases se realiza mediante una frontera de decisión elipsoidal [Müller et al., 2001].	87
5.5	El plano formado por el modelo de energía. Obtenido de [Li and Lu, 2009]	90
5.6	Validación cruzada de k pliegues, para K=4	93
5.7	Una curva ROC convencional, se muestran los tres puntos de operación posibles. Fuente [Metz, 1978].	95
6.1	Esquema con la extracción de características, entrenamiento y test del sistema basado en ARs para atributos continuos. . . .	102
6.2	Cortes coronales de la máscara 3D (basada en el promedio de controles) para: a)SPECT, b)FDG-PET(ADNI), c)PET-Cartuja	104
6.3	Corte coronal SPECT para mostrar el enmascarado basado en el promedio de normales en las fases de: a) SPECT original, b) promedio de controles, c) máscara, d) VAF: corte enmascarado	105

6.4	Definición de bloque tridimensional, con parámetro v para definir su tamaño y g para su solapamiento.	107
6.5	Diagrama de método de ARs sobre regiones discriminativas . .	110
6.6	Máscara 3D: a) promedio de controles para la base de datos SPECT, b) máscara FDR SPECT, c) máscara FDR para PET.	110
6.7	Esquema con las fases del sistema CAD propuesto: selección de características, entrenamiento y test.	113
6.8	Cortes transaxiales de la segmentación del histograma de imagen (SPECT en la primera fila y PET en la segunda) obtenidos de las imágenes promedio control (a) y g)) que se dividen respectivamente en las siguientes $K_1 = 5$ subimágenes correspondientes a cada intervalo.	114
6.9	Diagrama explicativo del sistema CAD de selección de características basado en ARs con el preprocesado, selección y extracción de características y clasificación.	116
6.10	Superficies de separación para los diferentes kernels de SVM al usar el método de FS basado en ARs y de FE clásico de PCA. Base de datos SPECT.	118
7.1	Extracción de características y diagrama de clasificación. Combinación de técnicas de reducción de características y clasificadores evaluados con validación cruzada de k -pliegues.	121
7.2	Corte coronal FDG-PET para mostrar el enmascarado en las fases de: a) original, b) promedio de controles, c) máscara, d) VAF: corte enmascarado	122
7.3	Magnitud de T para cada uno de los vóxeles dentro de la máscara y con un tamaño $5 \times 5 \times 5$	126
7.4	Superficies de separación para los diferentes “kernels” de SVM al usar el método de transformación LMNN junto con PCA como extractor de características. Base de datos SPECT.	129
7.5	Superficies de separación para los diferentes “kernels” de SVM al usar el método de reducción de características LMNN-RECT. Base de datos SPECT.	130
8.1	Cámara gamma para la obtención de SPECT del Hospital Universitario Virgen de las Nieves, Granada.	134
8.2	ADNI a nivel mundial: NA-ADNI (Norte América), Arg-ADNI (Argentina), E-ADNI (Europea), C-ADNI (China), K-ADNI (Coreana), J-ADNI (Japonesa), T-ADNI (Taiwanesa), A-ADNI (Australiana). Figura obtenida de [Weiner et al., 2012].	136

8.3	Regiones donde se detecta la evolución de la pérdida de memoria [Weiner et al., 2012]	138
8.4	Modelo para la progresión de AD. Obtenida de [Weiner et al., 2012]	143
9.1	SPECT. ROIs tridimensionales (centradas en los puntos) para un tamaño de voxel de 9, espacio de grid $g=4$ y activación por bloque de 90%	146
9.2	SPECT, método AR entre atributos continuos. ROIs activadas bidimensionales (cruces en los rectángulos) de los cortes axial, sagital y coronal obtenidos a partir de las ARs en modo supervisado con <i>minsup</i> y <i>minconf</i> de 100%. $t=90\%$	147
9.3	SPECT, método AR entre atributos continuos. Acc, Sen, Spe frente al porcentaje de reglas verificadas. Umbral de activación de bloque de 92%, con <i>minsup</i> y <i>minconf</i> de 100%.	148
9.4	SPECT. Regiones activadas bidimensionales (rectángulos) de los ejes axial, sagital y coronal con <i>minsup</i> y <i>minconf</i> de 90%.	149
9.5	Acc, Sen, Spe frente al % ARs verificadas (SPECT), $t=90\%$ para <i>minsup</i> , <i>minconf</i> :a) 100%, b) 95%	150
9.6	Acc, Sen, Spe frente al % ARs verificadas (SPECT), $t=90\%$ para <i>minsup</i> , <i>minconf</i> :a) 90% b) 80%.	151
9.7	Base de datos SPECT. Análisis ROC: método AR para atributos continuos con diferentes porcentajes seleccionados de <i>minsup</i> (100%, 95%, 90%, 80%) y <i>minconf</i> (100%, 95%, 90%, 80%) en la extracción de reglas. Comparación con otros métodos recientemente presentados que se representan por sus puntos de operación. Obsérvese el valor de pico de Acc =94.87%, con Sen= 91.07% y Spe= 100% (marcas en cruz), para <i>minsup</i> = <i>minconf</i> =100. La AUC obtenida para cada ROC es de: método AR (<i>minsup</i> : AUC) 100%: 0.9554, 95%: 0.9371, 90%: 0.9233, 80%: 0.9027; VAF SVM: 0.8993; PCA SVM: 0.9177	153
9.8	Base de datos PET Cartuja. ROIs activadas bidimensionales (cruces en los rectángulos) de los cortes axial, sagital y coronal obtenidos a partir de las ARs en modo supervisado con <i>minsup</i> y <i>minconf</i> de 100%).	155
9.9	Base de datos PET Cartuja. a) Acc, Spe y Sen en función del porcentaje de reglas verificadas (total de reglas=82082). b) Curva ROC en comparación con otros métodos publicados recientemente. AUC: VAF(0.7592), PCA(0.8683), AR(0.9158).	157

9.10	Comparación de las imágenes PIB-PET (rojo) con FDG-PET (verde). En amarillo se representa el área donde se solapan los mapas t de SPM [Shin et al., 2010].	158
9.11	Base de datos FDG-PIB. ARs pintadas como cuadrados en los cortes coronales para un $minsup, minconf=100\%$ para la base de datos PET:a) FDG, b) PIB, c) FDG más PIB (sobre una imagen media FDG de los controles). Grupo AD-NOR	158
9.12	Base de datos FDG-PIB, grupo MCI-NOR. ARs pintadas como cuadrados en los cortes coronales para un $minsup, minconf=100\%$ para la base de datos PET: a) FDG, b) PIB, c) FDG más PIB (sobre una imagen media FDG de los controles).	159
9.13	Base de datos FDG-PIB Exactitud (%) frente al porcentaje de reglas verificadas al utilizar FDG, PIB o combinación. a) CTRL-AD, b) MCI-NOR.	161
9.14	Base de datos FDG-PIB, método ARs entre atributos continuos. ROC del método propuesto en comparación con otros métodos publicados recientemente. a)NOR frente a AD b)ampliación del tramo superior.	162
9.15	Base de datos FDG-PIB, método ARs entre atributos continuos. ROC del método propuesto en comparación con otros métodos publicados recientemente. a) NOR frente a MCI b) ampliación del tramo superior.	163
9.16	Regiones activadas bidimensionales de los cortes coronales, axiales y sagitales para ARs entre atributos discriminantes. En la segunda fila se realiza una ampliación de los cortes cerebrales de la primera fila para representar las regiones cerebrales discriminativas de las que se extraen las ARs.	165
9.17	Valores de Acc, Sen, Spe (método ARs entre atributos discriminantes) como una función del umbral de reglas verificadas en (%) para a)SPECT b)PET. Comparación con otros métodos publicados recientemente.	166
9.18	Curva ROC (método ARs entre atributos discriminantes) para a) SPECT y b) PET. Comparación con otros métodos recientemente publicados.	168
9.19	Promedio de características con los valores de la activación media para las bases de datos SPECT y PET.	171
9.20	Curva ROC AR-discretos para bases de datos a)SPECT y b)PET. Comparación con el AR para características continuas y otros métodos publicados como: VAF-SVM y PCA-SVM.	173

9.21	Fracción (%) de las ROIs seleccionadas mediante ARs frente a <i>minsup</i> y <i>minconf</i> (%) para diferentes parámetros: umbral de activación y <i>minsup</i> / <i>minconf</i>	174
9.22	Valores de Acc en función del número de componentes (Base de datos SPECT, ARs como FS): a) PCA y b) PLS. c) Varianza explicada (%) como una función del número de componentes PCA y PLS en un diagrama de barras (las líneas representan la varianza explicada acumulada.)	175
9.23	Vóxeles seleccionados para PET por el método de selección de reglas en los cortes axiales, sagitales y coronales. a) la primera fila con <i>minsup</i> , <i>minconf</i> =100%, b) la segunda fila amplía las regiones donde aparecen las ARs seleccionadas en a), c) tercera fila con <i>minsup</i> , <i>minconf</i> =97%, activación 40%, d) la cuarta fila amplía las regiones donde aparecen las ARs seleccionadas en c)	178
9.24	Valores de Acc función del número de componentes extraídas para la base de datos FDG-PET: a) PCA, and b) PLS. Método FS-AR.	179
9.25	Curva ROC para la base de datos SPECT a)PCA y b)PLS. Método FS-AR. Comparación con otros métodos recientemente publicados.	180
9.26	Curva ROC PET a) PCA y b)PLS. Comparación con otros métodos recientemente publicados. La AUC obtenida para la ROC basada en AR-FS para PCA/PLS es: 100% <i>minconf</i> y <i>minsup</i> con 40% activación (0.93441/0.9391), 44% activación (0.9107/0.9181) y 97% <i>minsup</i> y <i>minconf</i> con 40% activación (0.9107/0.9181). AUC (PCA/PLS) para otros métodos publicados es: wilcoxon (0.8662/0.8662) o t test (0.8968/0.9027).	181
10.1	Varianza explicada (%) frente a las componentes principales y PLS en diagrama de barras. Las líneas representan la varianza explicada acumulada (%) frente a las componentes principales y las componentes PLS.	185
10.2	Clasificación SVM: Acc, Spe y Sen (%) frente al número de características reducidas para SPECT. Técnicas de reducción de características a) PCA + Transformación LMNN, b) PLS + Transformación LMNN, c)LMNN-RECT	191
10.3	Clasificación LMNN (Euclidea, Mahalanobis y modelos basados en energía) para SPECT. Técnicas de reducción de características a)PCA, b)PLS	192

10.4 a) Kernel SVM para características PLS transformadas LMNN
 b) SVM lineal sobre características directas PCA reducidas a la mitad VAF obtenidas para SPECT 193

10.5 a) Clasificación SVM Acc, Spe y Sen (%) frente al número de características reducidas para la base de datos PET. Técnicas de reducción de características: PCA más transformación LMNN, PLS más transformación LMNN, LMNN-RECT, PCA y VAF b) Clasificación LMNN (euclídea, mahalanobis y modelos basados en energía) para PCA y PLS 194

10.6 Los métodos basados en LMNN clasificados con SVM (Transformación PCA+ Transformación LMNN, PLS+Transformación LMNN, LMNN-RECT). Comparación con otros métodos recientemente publicados representados por sus puntos de operación. a) Base de datos SPECT. AUC: PCA+LMNN (0.9411), PLS+LMNN (0.9424), LMNN-RECT (0.9076), VAF SVM (0.8993) y PCA SVM (0.9177). b)PET database. La AUC obtenida para cada ROC es: PCA+LMNN (0.9505), PLS+LMNN (0.9437), LMNN-RECT (0.9325), VAF SVM (0.8500) and PCA SVM (0.9006). 195

13.1 3D Mask: a) average of controls for SPECT database, b) FDR-mask for SPECT database, and c) FDR-mask for PET database.228

13.2 Bidimensional activated regions (rectangles) from axial, sagittal and coronal axis with *minsup* and *minconf* of 100%, $t=90\%$. SPECT database, AR-continuous-based method. 232

13.3 AR-continuous-based method. Accuracy, Sensitivity and Specificity versus the percentage of verified rules when the activation block-threshold is 90% for *minsup* and *minconf* of a) 100%, b) 95%, c) 90% and d) 80%. Note that the overall number of mined rules and the Maximum accuracy rates obtained are a) 29756 rules, 94.87%, b) 84396 rules, 89.69%, c) 146768 rules, 87.63% and d) 288190 rules, 85.57% 233

- 13.4 ROC Analysis (SPECT). AR-continuous-based method: AR method with different selected percentages of *minsup* (100%, 95%, 90%, 80%) and *minconf* (100%, 95%, 90%, 80%) in the rule extraction. Comparison to other recently reported methods represented by their operation points. Observe the peak value of the Acc =94.87%, with Sen= 91.07% and Spe= 100% (cross-shaped mark), for *minsup*=*minconf*=100. The AUC obtained for each ROC is: ARs method (*minsup*: AUC) 100%: 0.9554, 95%: 0.9371, 90%: 0.9233, 80%: 0.9027; VAF SVM: 0.8993; PCA SVM: 0.9177 234
- 13.5 AR-continuous-based method, PET- Cartuja database. a) Acc, Spe y Sen versus the percentage of verified rules (total AR number=82082). b) ROC curve in comparison with other reported methods. AUC: VAF(0.7592), PCA(0.8683), AR(0.9158). 237
- 13.6 ARs depicted (as squares) in coronal slices at *minsup*, *minconf*=100% for PET database: a) FDG, b) PIB, c) FDG plus PIB (over FDG mean image) for the CTRL versus AD set. AR-continuous-based method, biomarker database. 238
- 13.7 Accuracy results for the set CTRL versus MCI at 100% of *minconf* and *minsup* when FDG, PIB or the combination of both are used. a) CTRL-AD, b) MCI-NOR.AR-continuous-based method, biomarker database. 239
- 13.8 Bidimensional activated regions from coronal, axial and sagittal slices FDR-AR-based method. FDG-PET database. 240
- 13.9 ROC Curve for FDR-AR-based method a) SPECT and b) PET databases. Comparison to other recently reported methods. . 243
- 13.10 Transaxial slice of the image histogram segmentation (SPECT in the first row and PET in the second one) obtained from control average images (depicted in (a) and (g)) which are respectively divided in the following $K_1 = 5$ subimages corresponding to each bin. Discretization-AR-based method. 244
- 13.11 ROC Curve a) SPECT and b) PET databases. Comparison with AR-baseline for continuous features and other reported methods: VAF-SVM and PCA-SVM. 246
- 13.12 Fraction (%) of AR-based selected ROIs versus *minsup* and *minconf* (%) for different parameters: Activation threshold and *minconf*/*minsup*. SPECT database. FS-AR based method. 248

13.13 Accuracy, sensitivity and specificity values as a function of the number of components for PET database: a) PCA, and b) PLS. FS-AR based method. 249

13.14 ROC Curve for SPECT database a) PCA and b) PLS. Comparison to other recently reported methods. FS-AR based method. 251

13.15 ROC Curve for PET database a) PCA and b) PLS. Comparison to other recently reported methods. The AUC obtained for each AR-FS-based ROC (PCA/PLS) is: 100% minconf and minsup with 40% activation (0.93441/0.9391), 44% activation (0.9107/0.9181) and 97% minsupp and minconf with 40% activation (0.9107/0.9181). AUC (PCA/PLS) for other reported methods is: wilcoxon (0.8662/0.8662) or t test (0.8968/0.9027). 252

14.1 Variance Explained (%) versus PCA and PLS Components in bar diagram. Lines represent the accumulated Variance Explained (%) versus Principal Components and PLS Components. 260

14.2 LMNN classification (Euclidean, Mahalanobis and Energy-based models) for SPECT Feature reduction techniques: a) PCA, b) PLS 265

14.3 a) SVM classification: Accuracy, Specificity and Sensitivity (%) versus number of reduced features for PET database. Feature reduction techniques: PCA plus LMNN Transformation, PLS plus LMNN Transformation, LMNN-RECT, PCA and VAF b) LMNN classification (Euclidean, Mahalanobis and Energy-based models) for PCA and PLS features 266

14.4 ROC Analysis. LMNN-based methods SVM classified (PCA+LMNN Transformation, PLS+LMNN Transformation, LMNN-RECT). Comparison to other recently reported methods represented by their operation points. a) SPECT database. The AUC obtained for each ROC is: PCA+LMNN (0.9411), PLS+LMNN (0.9424), LMNN-RECT (0.9076), VAF SVM (0.8993) and PCA SVM (0.9177). b) PET database. The AUC obtained for each ROC is: PCA+LMNN (0.9505), PLS+LMNN (0.9437), LMNN-RECT (0.9325), VAF SVM (0.8500) and PCA SVM (0.9006). 267

Índice de tablas

3.1	Comparación entre los principales trazadores de SPECT cerebral. Obtenida de [Díaz and Haro, 2004].	31
3.2	Comparación entre PET y SPECT (obtenida con HMPAO). Extraída de [Díaz and Haro, 2004]	33
5.1	Ejemplo de las transacciones para el estudio de un problema de cesta de la compra. Fuente: [Tan et al., 2005]	64
5.2	Representación binaria de imágenes funcionales. Fuente [Tan et al., 2005]	67
6.1	Ejemplo de representación binaria de imágenes funcionales. Fuente [Tan et al., 2005]	108
8.1	Información demográfica de la base de datos SPECT. μ and σ representa la media de la población y la desviación estándar respectivamente.	135
8.2	Detalles demográficos de la base de datos FDG-PET: μ y σ se refieren a la media y la desviación estándar de la población respectivamente	140
9.1	Medidas estadísticas del rendimiento de los métodos de selección de características basados en reglas usando varias configuraciones (<i>minconf</i> , <i>minsup</i> y % reglas verificadas). VAF, GMM y PCA se utilizan como referencia. Los parámetros VAF: clasificador SVM lineal, parámetros GMM: $\sigma=6$ clasificador RBF-SVM con 8 componentes y parámetros PCA: $\sigma=6$ y clasificador RBF-SVM con 16 componentes.	154

9.2	Comparación de la máxima exactitud y número de ARs extraídas y obtenidas para las bases de datos SPECT y PET para diferentes umbrales de la máscara de selección basada en vóxeles FDR.	164
9.3	Exactitud (%) tomando la quinta subimagen como máscara del total ($K_1 = 5$) y combinando diferentes intervalos en la fase de discretización ($K_2 = 5$). SPECT, ARs entre atributos discretos.	169
9.4	Tasa de exactitud (%) tomando la 5 ^a subimagen ($K_1 = 5$) como máscara y combinando diferentes intervalos de la fase de discretización ($K_2 = 5$). FDG-PET, ARs entre atributos discretos.	172
9.5	Comparación de Acc/Sen/Spe y AUC para diferentes técnicas FE con reducción final de 6 características (SPECT)	176
10.1	Medidas estadísticas de rendimiento de las técnicas basadas en LMNN en comparación con otros métodos publicados para la base de datos SPECT. Clasificador SVM: Comparación de los métodos publicados en este trabajo con los puntos de operación de VAF, GMM y PCA. Parámetros de las técnicas basadas en LMNN: SVM lineal con 6 componentes. Parámetros VAF: clasificador SVM lineal, parámetros GMM: clasificador RBF-SVM con $\sigma=6$, parámetros PCA: clasificador RBF-SVM con 16 componentes. Clasificador LMNN con 6 componentes euclídeas, de Mahalanobis y basada en energía.	188
12.1	Demographic details of the SPECT dataset. AD 1 = mild perfusion deficit, AD 2 = moderate deficit, AD 3 = severe deficit. μ and σ stands for population mean and standard deviation respectively.	218
12.2	Demographic details of the PET dataset. μ and σ stands for population mean and standard deviation respectively.	220
13.1	AR-continuous-based method. Statistical measures of performance of AR-based feature selection methods for the sample group using several set-ups (<i>minconf</i> , <i>minsup</i> and % verified rules). Voxel-as-features (VAF), GMM and PCA operation results are reported as reference. VAF parameters: linear SVM classifier, GMM parameters: $\sigma=6$ RBF-SVM classifier with 8 components and PCA parameters: $\sigma=6$ RBF-SVM classifier with 16 components.	236

13.2 Comparison of maximum Accuracy and number of AR mined obtained for SPECT and PET databases for different thresholds of the FDR voxel selection mask. 241

13.3 Accuracy rate (%) taking 4th bin subimage ($K_1 = 5$) as mask and combining different bins of discretization stage ($K_2 = 5$) for SPECT/PET databases. 245

13.4 Comparison of Acc/Sens/Spe rate and AUC for different FE techniques with final reduction of 6 features (SPECT database. FS-AR based method.) 247

14.1 Statistical measures of performance of LMNN-based techniques in comparison with other reported methods for SPECT database. SVM classifier: Comparison of the methods reported in this work with VAF, GMM and PCA operation points. LMNN-based techniques parameters: linear SVM classifier with 6 components. VAF parameters: linear SVM classifier, GMM parameters: $\sigma=6$ RBF-SVM classifier with 8 components and PCA parameters: $\sigma=6$ RBF-SVM classifier with 16 components. LMNN Classifier with 6 components: Euclidean, Mahalanobis and Energy. 261

Parte I

Introducción

Capítulo 1

Introducción

En este capítulo introductorio pretendemos, por un lado, poner de manifiesto la necesidad de desarrollar sistemas de ayuda al diagnóstico de alteraciones neurológicas. En este sentido, la principal motivación reside en las grandes ventajas que otorga un diagnóstico en etapas iniciales de la enfermedad, por ejemplo en el caso de la Enfermedad de Alzheimer (del inglés “Alzheimer’s Disease”, AD), permite que el paciente pueda participar en la planificación de su vida, así como recibir tratamiento que ralentice el devenir de esta enfermedad. De esta forma, ofrecemos un pequeño resumen del estado del arte de los llamados sistemas CAD (del inglés “computer-aided diagnosis”) basados en tomografía computerizada, que ayudan al médico especialista a establecer un diagnóstico precoz. Estos sistemas CAD que conforman el estado del arte serán analizados en detalle en siguientes capítulos. Por otro lado, en este capítulo establecemos los objetivos de la presente Tesis, así como la estructura de contenidos de la misma. Dentro de los objetivos destacamos como principal, el desarrollo de un sistema CAD automático basado en reglas de asociación (AR, del inglés “association rules”) sobre regiones de interés en imágenes de tomografía funcional. Las ARs se han empleado en otros contextos como el análisis de mercados, no habiéndose explorado su funcionamiento en este campo. Finalmente presentamos la organización de los contenidos de la memoria de manera detallada.

1.1 Motivación

Al envejecer aumenta el riesgo de desarrollar una enfermedad neurológica, siendo la enfermedad de Alzheimer (AD, del inglés “Alzheimer’s Disease”) [Swerdlow, 2011] la más frecuente hoy en día. En la última década se han producido importantes avances en el reconocimiento de los factores genéticos [Schellenberg et al., 1989] y del estilo de vida de cada individuo [Cappa et al., 2004] que puedan relacionarse con la aparición de la misma, pero aún estamos lejos de entender las verdaderas causas [Giasson et al., 2002] para favorecer un diagnóstico precoz a tiempo [Chincarini et al., 2011], [Feldman, 2007].

La AD es un desorden neurodegenerativo que afecta primero a las funciones de memoria y gradualmente a todas las funciones cognitivas con desajustes de comportamiento, causando finalmente la muerte [Henley et al., 2012]. La AD constituye la principal causa de demencia en la tercera edad, afecta a 30 millones de personas mundialmente y se espera que esta cifra se triplique en los próximos 50 años, lo que no sólo supone un problema social, sino un reto socioeconómico para los países desarrollados y cada vez más, para aquellos en vías de desarrollo. Como dato destacable, en 2050, se espera que el número de personas mayores en los países desarrollados sea cuatro veces el de 1950 (95 millones frente a 416 millones) y que el número de niños descienda de 222 millones en 1950 a 197 millones en 2050 [Nations, 2008].

Dado que la AD no tiene cura, cada vez son más necesarias las medidas que no sólo ayuden a retrasar su desarrollo, sino a poder prevenirla para poner en marcha las medidas terapéuticas tan pronto como sea posible y así aumentar la efectividad del tratamiento [Braak and Tredici, 2012]. El diagnóstico de la AD está basado en la información provista por un examen clínico cuidadoso, una detallada entrevista al paciente y a sus familiares junto con una valoración neuropsicológica [Ewers et al., 2012]. Sin embargo, distinguir AD usando estos protocolos clínicos (o test neuropsicológicos) constituye un reto de diagnóstico especialmente durante las etapas tempranas de la enfermedad. Con la misma motivación que los tests neuropsicológicos, se ha demostrado que las imágenes funcionales cerebrales proporcionadas por la tomografía computarizada de emisión de fotón único (SPECT, del inglés “Single Photon Emission Tomography”) y la tomografía de emisión de positrones (PET, del inglés “Positron Emission Tomography”) [Ito, 2006] son muy efectivas para la extracción de datos en el diagnóstico temprano de la AD usando métodos no invasivos [Ramírez et al., 2009b], [Salas-Gonzalez et al., 2010b], [Illán et al., 2010]. En el contexto del año de la investigación de Alzheimer [Conesa, 2011], se han dado a conocer numerosas iniciativas que tienen ese propósito, si bien en su mayor parte se basan en desarrollar sistemas de diagnóstico por imagen

[Chincarini et al., 2011], ya sea por resonancia magnética [Tondelli et al., 2012] o mediante la tomografía de emisión de positrones y/o fotones [López et al., 2009a], [Górriz et al., 2011].

1.2 Estado del arte

Los distintivos patológicos de la AD son las placas $A\beta$ amiloide, ovillos neurofibrilares y gliosis reactiva. El diagnóstico actual de la AD se lleva a cabo mediante la valoración clínica, neuropsicológica y de neuroimagen [Villemagne et al., 2005]. Los criterios para el diagnóstico clínico de la AD fueron establecidos por el Instituto Nacional de Desórdenes Neurológicos y Comunicativos (NINCDS) junto con la asociación de la AD y desórdenes relacionados (ADRDA) en [Jack et al., 2011]. El problema es que al tratarse de criterios clínicos, cuando se cumplen, los pacientes ya han desarrollado la enfermedad y por tanto no son útiles para realizar un diagnóstico precoz de la misma.

Las técnicas de neuroimagen ofrecen información complementaria para mejorar la precisión en el diagnóstico de la demencia y se pueden clasificar en: neuroimagen estructural y funcional. La primera comprende la tomografía computerizada (CT, del inglés “Computed Tomography”), y la resonancia magnética (MRI, del inglés “Magnetic Resonance Image”). La segunda agrupa a la espectrografía de resonancia magnética (MRS, del inglés “Magnetic Resonance Spectrography”), SPECT, PET y resonancia magnética funcional (fMRI, del inglés “functional Magnetic Resonance Imaging”). La imagen funcional tiene la capacidad de identificar cambios patopsicológicos en el cerebro antes que la imagen estructural, favoreciendo en este sentido el diagnóstico precoz de la AD [Villemagne et al., 2005].

El estudio de las técnicas de neuroimagen está relacionado con el biomarcador analizado. La mayoría de la investigación sobre el Alzheimer llevada a cabo entre 2000 y 2010 se ha centrado en encontrar biomarcadores que podrían ser usados de forma fiable para diagnosticar la AD, monitorizar su progresión y predecir su aparición [Caroli et al., 2010]. Los biomarcadores son parámetros (fisiológicos, bioquímicos, anatómicos o de imagen médica) que pueden medirse *in vivo* y que reflejan características específicas de la enfermedad relacionadas con procesos patopsicológicos. Algunos biomarcadores de la AD son $A\beta$ 1-42 [Weiner et al., 2012] y la concentración de la proteína tau del fluido cefalorraquídeo (CSF, del inglés “Cerebrospinal Fluid”) [Rubio et al., 1998], volumen del hipocampo en MRI [Chincarini et al., 2011], fluorodeoxiglucosa (FDG, del inglés “Fluorodeoxyglucose”) y/o componente de Pittsburg (C-PIB, del inglés “Pittsburg-B Compund”) [Shin et al., 2010] en PET. La disponibilidad de

distintos biomarcadores validados abre la discusión de cómo elegirlos, es decir cuál predice mejor AD, así como el hecho de combinar la información de varios de ellos [Illán et al., 2010].

1.2.1 Sistemas de ayuda al diagnóstico

El diagnóstico asistido por computador (CAD, del inglés “Computer Aided Diagnosis”) [Fujita et al., 2008] aporta mediante técnicas automáticas de análisis de datos una *segunda opinión* que sirve de ayuda a la toma de decisiones diagnósticas por parte de los médicos especialistas. Lo que se pretende con los sistemas CAD no es que tengan una precisión igual o superior a la de los médicos sino que proporcionen información complementaria de ayuda en la toma de decisiones [Segovia et al., 2010], [Salas-Gonzalez et al., 2008], [Chaves et al., 2009b].

Los sistemas CAD para el tratamiento de imágenes se pueden clasificar, según la forma en que procesan la información, en univariados y multivariados. El enfoque univariado procesa las imágenes vóxel a vóxel y la aproximación más relevante está basada en mapas estadísticos paramétricos (SPM, del inglés “Statistical Parametric Mapping”) [Ashburner, 2000], [Friston et al., 1994] que compara los valores del vóxel de la imagen bajo estudio con un grupo de pacientes normales. Posteriormente, los vóxeles relevantes de este test son inferidos usando la teoría de campos aleatorios. SPM se diseñó para la comparación de conjuntos de imágenes y no para el diagnóstico automático del paciente bajo estudio (asunto resuelto por los modelos multivariados).

En las aproximaciones multivariadas, a pesar de que los datos para la clasificación se determinan a partir de la imagen cerebral completa, se utiliza sólo un conjunto reducido de vóxeles o regiones para distinguir entre imágenes patológicas y controles [Illán et al., 2011a], [Friston et al., 2007]. La clasificación se lleva a cabo mediante la definición de vectores de características y entrenamiento de un clasificador asumiendo que las hipótesis de partida de las etiquetas de entrenamiento y test son válidas por lo que una precisión elevada en la clasificación es equivalente a un diagnóstico efectivo del paciente bajo estudio. El auge de los sistemas multivariados se debe, en gran parte, a los avances recientes en la definición de los clasificadores estadísticos para construir clasificadores más fiables y con mayor capacidad de generalización, capaces de afrontar mejor el problema del pequeño tamaño muestral [Duin, 2000]. Algunas aproximaciones que aplican CAD multivariados se recogen en [Illán et al., 2011a], [Ramírez et al., 2009b], [Segovia et al., 2010], [Salas-Gonzalez et al., 2008] etc.

El grupo de investigación en Procesamiento de señal y aplicaciones biomédicas

(SIPBA, del inglés “Signal Processing and Biomedical Applications”) de la Universidad de Granada ha llevado a cabo importantes avances en el estudio de la detección precoz de alteraciones neurológicas mediante el desarrollo de técnicas de preprocesamiento, análisis y clasificación de imágenes biomédicas PET y SPECT. En [Salas-Gonzalez, 2009] se presentan métodos de optimización mejorados para la normalización afín basados en el algoritmo de Levenberg-Marquardt y de Gauss-Newton incluyendo un parámetro adicional para optimizar la búsqueda de la solución. También se plantean novedosos modelos de selección de vóxeles usando la imagen media, la desviación estándar y el test de Mann-Wilcoxon, así como un método para calcular las regiones de interés usando análisis factorial. En el trabajo descrito en [I.A.Illán, 2009] se profundiza en el análisis de componentes (principales: PCA del inglés “Principal Component Analysis”, independientes: ICA del inglés “Independent Component Analysis” o mediante técnicas de factorización) como extractores de características. Por otro lado, se incide en el análisis de componentes principales (mediante la transformación de Karhunen-Loève y la selección de eigenbrains mediante el criterio de Fisher y PCA), en el análisis discriminante lineal (LDA, del inglés “Lineal Discriminant Analysis”) combinado (o no) con PCA en [Pérez, 2010]. También se profundiza en el desarrollo de métodos de aprendizaje basados en núcleo (“kernel” en inglés) como son las máquinas de vectores de soporte (SVMs, del inglés “Support Vector Machines”), Kernel PCA (KPCA) y en el análisis discriminante kernelizado (KDA, del inglés “Kernel Discriminant Analysis”). Por último, en [F.Segovia, 2010] se proponen métodos de agrupación (“clustering” en inglés) basados en mezcla de Gaussianas. Su objetivo es construir super-vóxeles que definan regiones de interés para posterior clasificación. Adicionalmente en la anterior referencia se introdujo la técnica de los mínimos cuadrados parciales (PLS, del inglés “Partial Least Squares”) como método de extracción de características para imagen funcional PET y SPECT, obteniendo mejoras significativas respecto a técnicas clásicas de extracción como PCA.

La principal aportación que lleva a cabo la presente tesis doctoral respecto a las otras previamente citadas es el estudio del cerebro en términos de interconectividad de imágenes funcionales mediante ARs, siendo esta una forma dinámica y novedosa de analizar enfermedades neurodegenerativas como las que nos ocupa en este trabajo. Las ARs se utilizan como [Chaves et al., 2012e] selector de características y clasificador según el número de reglas verificadas por cada paciente tanto en atributos continuos [Chaves et al., 2011a], discretos [Chaves et al., 2013] o discriminantes [Chaves et al., 2012c] con una experimentación llevada a cabo en diferentes bases de datos para demostrar la ro-

bustez del método así como la aportación de diferentes biomarcadores [Chaves et al., 2012a]. Asimismo se analiza el algoritmo LMNN [Chaves et al., 2012d] como transformador del espacio de características que consigue discriminar mejor las clases bajo estudio (Control frente a AD) y proporcionar una tasa de clasificación estable para un amplio rango de características; como reductor del espacio de características mediante una matriz rectangular (LMNN-RECT) y/o como clasificador.

1.3 Objetivos

Los **objetivos** de este trabajo se pueden resumir en:

- Con la motivación del diagnóstico precoz de la AD, se pretende desarrollar un sistema CAD con nuevas técnicas de aprendizaje supervisado para la extracción de características y clasificación basadas en ARs y LMNN.
- Mostrar la generalidad de los métodos basados en ARs y LMNN usando diferentes bases de datos para la experimentación, estudiando diferentes técnicas de imagen y resaltando la robustez de los algoritmos presentados.
- Combinación y estudio comparativo de distintos biomarcadores (FDG y C-PIB) para favorecer la detección de la enfermedad en sus estadios más tempranos.
- Profundizar en el diseño de algoritmos eficientes, mediante diferentes técnicas de enmascaramiento que favorezcan la focalización en las regiones más discriminativas.
- Profundizar en la selección y clasificación de características de atributo continuo mediante ARs y comparar con el concepto de atributo discreto.

1.4 Organización de la memoria

En esta sección, presentamos la organización de esta tesis. La **primera parte**; incluye un primer capítulo introductorio para comentar las motivaciones de este trabajo, el estado del arte en el análisis de imagen médica y su aplicación a enfermedades neurodegenerativas como la AD, así como los objetivos y la organización de la memoria. El segundo capítulo recoge los contenidos relativos a la AD, por ejemplo los factores genéticos que la determinan, los biomarcadores

que facilitan su detección, nociones de histopatología, tests neuropsicológicos para su diagnóstico, opciones de tratamiento y futuro de la AD.

La **segunda parte** presenta el estado del arte con la descripción de las técnicas tomográficas usadas en medicina nuclear en el capítulo 3. Este capítulo se centra además en las técnicas de imagen funcional utilizadas en esta tesis, es decir, SPECT y PET. Se comenta la forma de adquisición de las mismas así como su reconstrucción mediante retroproyección filtrada. También se incide en la normalización (espacial y en intensidad), que es una etapa fundamental para que los vóxeles de todas las imágenes analizadas se correspondan en el espacio cerebral con exactitud y su intensidad sea comparable entre diferentes observaciones. El capítulo 4 incide en el diseño asistido por computador para la detección precoz de enfermedades neurodegenerativas, en sus dos aproximaciones más conocidas para el procesado de las imágenes: univariadas, con especial incidencia en los mapas estadísticos paramétricos (SPM) y multivariadas.

La **tercera parte** (capítulo 5) profundiza en los fundamentos teóricos utilizados en este trabajo. Se presentan en primer lugar los métodos de selección de características clasificados en tipo filtro, envoltorio y/o embebidos para enlazar con las ARs que también se diseñan en este trabajo como selectores, opción descrita en la sección 5.1 de análisis de imagen funcional. También se explica el criterio del ratio discriminante de Fisher (FDR, del inglés Fisher Discriminant Ratio) para la selección de atributos discriminativos entre los que se establecerán ARs. Se describe en segundo lugar el problema de las ARs así como los algoritmos más utilizados para su extracción, con especial reseña en el algoritmo *Apriori* (el empleado en esta tesis) y la aplicación de las ARs a imagen médica.

Del mismo modo se presentan diferentes tipos de extractores de características de las previamente seleccionadas como entrada eficaz y reducida a un clasificador. De esta forma proponemos nuevos métodos de extracción regidos por distancias y basados en el algoritmo LMNN que pretenden minimizar una función de pérdidas para ampliar el margen entre clases diferentes. A nivel teórico detallamos las opciones de kernel LMNN y la transformación LMNN y LMNN-RECT para transformar o reducir el espacio de características. Por último, se presentan métodos de validación estadística para evaluar la *bondad* del clasificador como dejar uno fuera (LOO, del inglés “leave-one-out”) o k-pliegues (del inglés “k-fold”). Se establecen nociones teóricas de clasificación profundizando en las máquinas de vectores de soporte, en el clasificador LMNN y el basado en reglas, los cuales serán empleados en las secciones 10.3 y 10.4 de experimentación del trabajo para comprobar el rendimiento de los

métodos propuestos.

La **cuarta parte** presenta los métodos de análisis de imagen funcional propuestos en esta tesis. El capítulo 6 se centra en las ARs aplicadas a imagen médica para la detección de AD, que se establecen entre atributos continuos, discriminativos o discretos y se lleva a cabo la clasificación mediante un sistema ideado basado en el número de reglas verificadas por cada sujeto. También se presentan las ARs como un método eficaz de selección de características.

El capítulo 7 se centra en el algoritmo LMNN, que tiene la finalidad de aproximar los vecinos con la misma etiqueta y separar los que tengan etiqueta diferente. LMNN se puede utilizar como transformador de características previamente seleccionadas así como reductor de características. En ambos casos, la clasificación se realiza con kernel SVM. Por otro lado, LMNN también se usa como clasificador con modelos basados en energía o con distancia de Mahalanobis, mejorando ambas la clasificación mediante distancia Euclídea.

En la **quinta parte** se muestran los resultados experimentales. El capítulo 8 se destina a la descripción de las bases de datos utilizadas. Las bases de datos utilizadas se clasifican en tomografía de emisión de fotón único (SPECT) y de positrones (PET) según el radioisótopo usado. Respecto a la segunda modalidad (PET) se dispone de diferentes bases de datos según los biomarcadores de imagen utilizados, por ejemplo la fluorodeoxiglucosa (FDG) o el componente de Pittsburg (C-PIB), con resultados complementarios como se demuestra en la sección 9.1 de experimentación de esta tesis.

La procedencia de los escáneres es diversa, desde el Hospital Universitario Virgen de las Nieves en Granada, o el Hospital de Cartuja de Sevilla (España) hasta la iniciativa de neuroimagen de la enfermedad de Alzheimer (ADNI, del inglés “Alzheimer’s Disease Neuroimaging Initiative”) en California (USA).

En el capítulo 9 se muestran los resultados experimentales con ARs y en el capítulo 10 se lleva a cabo el análisis experimental con LMNN según los métodos de análisis de imagen descritos en la sección previa. Para comprobar la robustez del sistema se usaron las diferentes bases de datos descritas en el capítulo 8. Se llevaron medidas estadísticas de exactitud, especificidad y sensibilidad, estudios mediante curvas características operativas de receptor (ROC, del inglés “Receiver Operating Characteristic”) y análisis del área bajo la curva (AUC, del inglés “Area Under Curve”) para mostrar la efectividad de cada método y compararlo con otros recientemente publicados.

La **sexta parte** (capítulo 11) muestra las conclusiones y posibilidades de trabajo futuro en las líneas desarrolladas. No sólo se aprecia que con los métodos presentados, la exactitud obtenida supera siempre el 90%, sino que también se incide en la novedad de los algoritmos presentados en términos

de análisis de neuroimagen a día de hoy. No obstante y aunque este trabajo engloba la solución a un problema de detección precoz de AD desde diferentes puntos de vista, son líneas de investigación abiertas, por lo que se comentan opciones de mejora.

En la **séptima parte** (capítulos 12 a 15) se lleva a cabo un resumen en inglés de la memoria para favorecer la difusión del trabajo presentado y cumplir con los requerimientos para la obtención del doctorado internacional.

Capítulo 2

La Enfermedad de Alzheimer

Este capítulo define las características de la AD, haciendo referencia primero a la incidencia en la salud que está adquiriendo esta patología a nivel mundial junto con sus repercusiones económicas. Establecemos después las características histopatológicas para diferenciar la AD de otras demencias y se fundamenta su correlación con la genética. También se presentan los biomarcadores que pueden reflejar potencialmente los cambios bioquímicos, genéticos y/o las características extraídas de las imágenes médicas para la detección precoz de la AD, y se clasifican respectivamente en: biomarcadores en sangre y/o CSF, genéticos y de imagen. Por último se comentan las opciones para el tratamiento de la AD, junto con el futuro que se predice de la enfermedad.

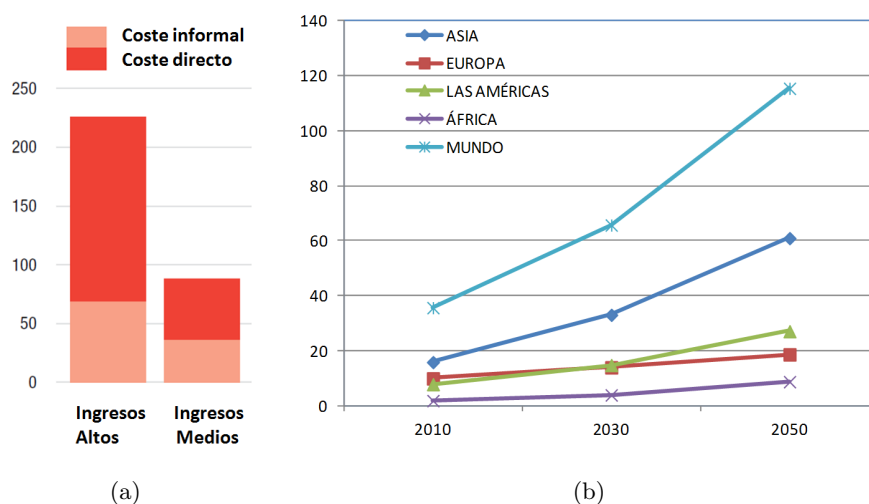


Figura 2.1: a) Costes directos e indirectos en billones de dólares en países de altos o medios ingresos, b) Número de personas con demencia en millones frente a la evolución en años desde 2010 a 2050. Fuente: [Nations, 2008]

2.1 Introducción

La AD es la causa más común de demencia en la tercera edad y está caracterizada por un espectro de características clínicas y fallos neuropatológicos [Bennett and Evans, 1992] que aparecen progresivamente. A comienzos del siglo 20, el neurólogo alemán Max Bielchowsky visualizó componentes celulares dentro de las neuronas que denominó ovillos neurofibrilares. En 1906, utilizando el método de Bielchowsky, el Dr. Alois Alzheimer describió el comienzo de una nueva patología en el cerebro de una de sus pacientes: Auguste D., de 51 años. La paciente presentaba una serie de síntomas clínicos incluyendo desorden amnésico progresivo, afasia, alexia y agrafia, acompañados de desorientación, comportamiento impredecible, agitación profunda, alucinación auditiva, paranoia y una pronunciada discapacidad psicosocial [Feldman, 2007]. La rapidez de degeneración y la relativa juventud de la paciente motivaron al Dr. Alzheimer a investigar sus características neuropatológicas [Alzheimer, 1907]. De un cuarto a un tercio de las neuronas cerebrales corticales habían desaparecido, y muchas de las neuronas contenían gruesas fibrillas enrolladas dentro de su citoplasma. De hecho, el Dr. Alzheimer sugirió que los ovillos neurofibrilares ocasionaban la muerte celular, dejando la neurofibrilla como marcador.

A partir de 1960 se realizaron estudios más sofisticados que mostraban que la estructura de un ovillo neurofibrilar contiene masas de fibras con una estructura periódica denominada filamento helicoidal pareado (del inglés “Paired Helical Filaments”, PHF) [Thal et al., 1998]. Estos descubrimientos permitieron definir la estructura del núcleo de amiloide (placa neurítica) para desarrollar métodos de aislamiento de placas, ovillos neurofibrilares y PHFs (ver más detalles en las secciones de *Genética e Histopatología de la AD* de este capítulo).

Aunque han existido numerosos y productivos estudios de investigación de placas y ovillos neurofibrilares, la motivación central concierne a la validez de la AD como una categoría diferente de la demencia senil [Nakano et al., 1986], así como si la AD es realmente la causa de la demencia en la tercera edad, y si la senectud normal o la arterioesclerosis cerebral son las responsables de la discapacidad cognitiva en las personas mayores. De hecho, hay teorías que defienden que la AD es una consecuencia del envejecimiento cerebral y otras que no. Actualmente se postula desde la perspectiva de la mitocondria, orgánulo que parece jugar un papel decisivo tanto en la AD como en el envejecimiento cerebral. La hipótesis en cascada de la mitocondria postula que hay mecanismos comunes que conducen al envejecimiento cerebral y a la AD, así como que la producción de placas, ovillos neurofibrilares y la degeneración sináptica son consecuencias de la funcionalidad perturbada de la mitocondria [Swerdlow, 2011].

En la actualidad, la AD es uno de los desórdenes neurodegenerativos más severos y frecuentes en la población de la tercera edad, y tiene implicaciones importantes y dramáticas tanto de la salud como socioeconómicas. Debido al incremento de la esperanza de vida, se espera que la incidencia de la AD se duplique durante los próximos 20 años principalmente en los Estados Unidos, Europa y Japón [Mueller et al., 2005]. Así la figura 2.1.a muestra los costes directos e informales de la demencia en países con altos o medios ingresos, observándose que los primeros (entre los que se encuentran Estados Unidos, Europa y Japón) invierten más del doble respecto a los otros. La figura 2.1.b muestra la evolución del número de pacientes con demencia desde 2010 a 2050 en diferentes países. Se observa que, dada la tendencia de la gráfica, en 2050 el número de pacientes con AD se habrá aproximadamente triplicado mundialmente respecto a 2010.

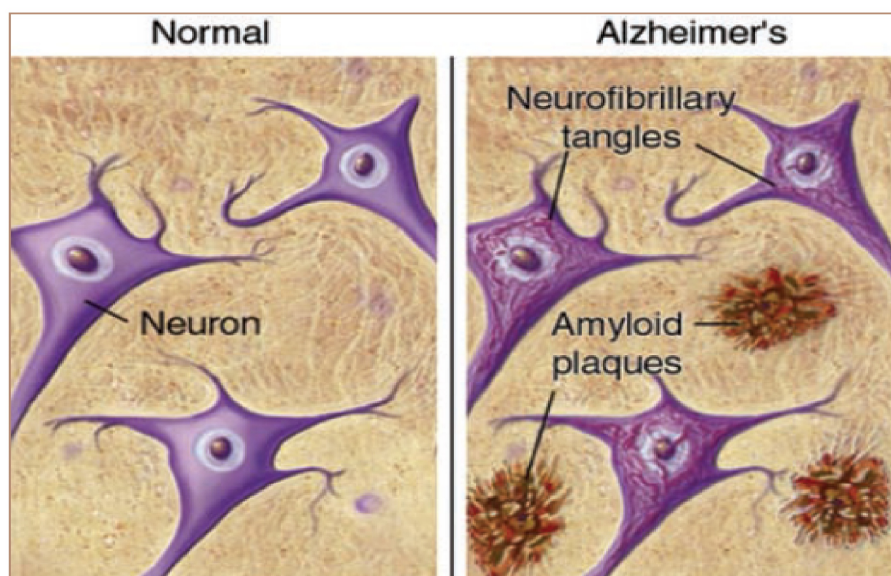


Figura 2.2: Ovillos neurofibrilares y placas seniles de los pacientes con AD respecto a los normales. Figura obtenida de [Feldman, 2007].

2.2 Histopatología de la AD

Desde una perspectiva patológica, las placas neuríticas y los ovillos neurofibrilares sirven como línea divisoria entre la AD y otras demencias. Hay dos marcas neuropatológicas fundamentales en la AD (como se muestra en la figura 2.2): la placa senil neurítica, que contiene proteína beta amiloide ($A\beta$) y se deriva de procesar una proteína mayor denominada proteína precursora de amiloide (APP, del inglés “Amyloid Precursor Protein”), y el ovillo neurofibrilar, un filamento helicoidal pareado con proteína tau hiperfosforilada [Martin et al., 2011]. Los ovillos neurofibrilares son colecciones intracelulares de filamentos anormales que desplazan los orgánulos citoplasmáticos normales. En la AD, los ovillos neurofibrilares tienden a ser más numerosas en las estructuras límbicas del lóbulo temporal mesial, incluyendo el hipocampo, sibilum, amígdala, corteza entorrinal y transentorrinal. Por ejemplo, en las neuronas piramidales, tales como las del hipocampo, los ovillos neurofibrilares tienden a rellenar gran parte del cuerpo celular y de las dendritas persistiendo incluso en el espacio extracelular después que la neurona muera.

Estas lesiones neuropatológicas empiezan a formarse probablemente años antes de la completa expresión de la demencia clínica, particularmente en los

estadios de deterioro cognitivo leve (MCI, del inglés “Mild Cognitive Impairment”) que son a menudo prodromales a la AD [Gillain et al., 2009]. Bajo el depósito de placas seniles y de ovillos neurofibrilares hay un estrés resultante de inflamación y oxidación que se añaden a la cadena patológica de consecuencias. Como resultado, hay una pérdida de neurotransmisores y de neurotransmisión que ayudan a comprender en mayor profundidad el progreso de la AD para definir mejor los objetivos de tratamiento. Aunque las placas seniles y los ovillos neurofibrilares se consideran cambios patológicos de la AD, no son específicos. De hecho, no está claro qué proporción de estos casos representa la AD preclínica frente a un cambio limitado relacionado con la edad que no ha progresado a AD.

2.3 Genética de la AD

En 1984, Glener y Wong informaron de la secuencia de amino ácidos de la componente principal de beta amiloide, un péptido 4-kDa denominado proteína beta amiloide ($A\beta$). El descubrimiento de mutaciones dominantes autosómicas en la APP anunció numerosos adelantos en el estudio de la genética de la AD [Findeis, 2007]. Durante este período de descubrimiento molecular, se identificaron además factores de riesgo genéticos en formas familiares y esporádicas de la AD. Se demostró por ejemplo que los polimorfismos del gen de la apolipoproteína E (APOE) estaban involucrados en los casos de efectos tardíos de la AD ya que ésta puede involucrar $A\beta$ y localizarse en placas seniles.

Desde la identificación inicial de la $A\beta$ como la componente principal de las placas de amiloide, se ha evidenciado que la acumulación de este péptido incita sucesos neuropatológicos en la AD. De acuerdo con esta hipótesis está la observación de que la amplia mayoría de las mutaciones causantes de los primeros efectos de la AD incrementa el ratio de la fibrilogénica $A\beta$ 42. Así la figura 2.3 muestra relativo a una edad fija (65 años), el depósito de $A\beta$ en rojo, el daño neurodegenerativo en azul y medidas estructurales MRI de la atrofia cerebral en verde para el gen APOE ϵ 4 no portador de AD (2.3.a) y portador de AD(2.3.b). Se demuestra que para el caso b (portador), el efecto hipótesis consiste en que se almacenan las placas de amiloide precozmente respecto al caso a, de lo que se deduce que los sujetos portadores del gen APOE ϵ 4 tienen el riesgo de desarrollar la AD a una edad más temprana que los no portadores. Se destaca que el depósito de amiloide frente a los otros dos biomarcadores se manifiesta con anterioridad. En general, hay 3 genes asociados con la familia de AD temprano: APP, presenilin-1 (PS-1) y presenilin-2 (PS-2) que se sitúan

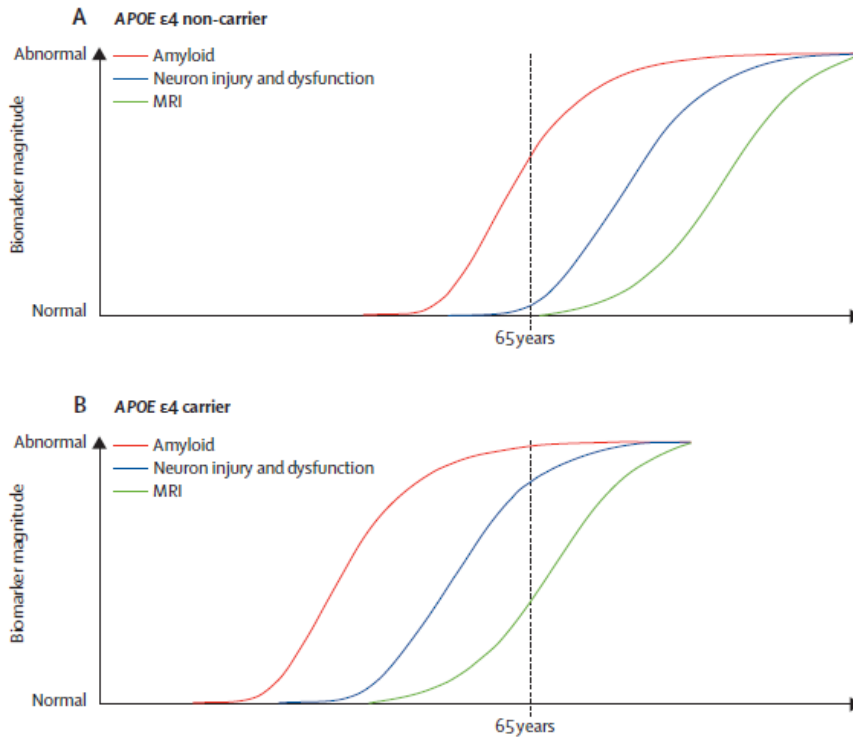


Figura 2.3: Moduladores de las relaciones temporales de biomarcador relativos a una edad fija (65 años). Hipótesis APOE $\epsilon 4$ portador (A) y no portador (B). Extraída de [Jack et al., 2010].

en los genes 21, 14 y 1 respectivamente [Wolfe, 2002]. Las mutaciones de estos genes, heredadas de una forma dominante autosómica, causan la AD.

2.4 Biomarcadores para la AD

En la actualidad no existen biomarcadores disponibles para el diagnóstico o la monitorización de la evolución de la AD. Esto ha sido un gran impedimento en la investigación de las terapias que afectan a la enfermedad, aunque puede ser previsiblemente resuelto ya que existe un gran interés a día de hoy en la medida de la concentración de proteína tau en el CSF y del $A\beta$ en sus variadas formas. Dado que el CSF está en contacto directo con el espacio extracelular del sistema nervioso central, puede reflejar potencialmente los cambios bioquímicos en el cerebro [Hampel et al., 2003]. Respecto a los marcadores de CSF, a través

de la monitorización de la proteína tau en el CSF y del β -amiloide, puede ser posible controlar la huella neuropatológica de la AD. De hecho, numerosas investigaciones se han basado en el hecho de haber encontrado bajos niveles de $A\beta$ 1-42 y altos niveles de proteína tau total en el CSF (T-tau) tanto en AD temprana como en MCI que posteriormente se convirtieron a AD. Por otro lado, los elevados niveles de proteína tau en el CSF [Martin et al., 2011] en las etapas más tardías de la AD predicen una neurodegeneración severa. El cómo los niveles de los biomarcadores CSF reflejan la severidad de los síntomas o la progresión de la AD está aún investigándose. Sin embargo, es importante reconocer que los biomarcadores actuales de CSF en aislamiento son insuficientes para diagnosticar o excluir posibles casos de AD. Es necesario añadir datos clínicos, neuropsicológicos, de laboratorio y de neuroimagen.

Respecto a los biomarcadores de la sangre, se destaca su accesibilidad y monitorización para detección de la AD, la facilidad de obtención y el menor número de efectos laterales que el testeo del CSF. De hecho, actualmente hay un interés creciente en la medición de isoprostanos en la sangre y CSF [Hampel et al., 2003]. Los isoprostanos son compuestos parecidos a la prostaglandina que se producen por los radicales libres mediados por la peroxidación de las lipoproteínas, un proceso que puede ser perjudicial para las neuronas. Los marcadores de la inflamación (por ejemplo: la proteína C-reactiva, el factor de necrosis tumoral o la interleucina) y del metabolismo (como por ejemplo la encima que degrada la insulina, la hormona del crecimiento, el cortisol y otros esteroides) degradan las encimas, el crecimiento hormonal, el cortisol y otros esteroides de lo que se han elaborado informes por su correlación con los descubrimientos clínicos de la edad y la disminución cognitiva.

Aparte de los biomarcadores en sangre y/o CSF [Hampel et al., 2003] y los biomarcadores genéticos [Wolfe, 2002] (explicados en la sección anterior), existen los llamados biomarcadores de imagen que definen características objetivas extraídas de las imágenes médicas, relacionadas con procesos biológicos normales, enfermedades o respuestas terapéuticas.

En la práctica clínica, estos biomarcadores pueden ser de gran interés por las ventajas que proporcionan al proceso de diagnóstico, terapéutico y de seguimiento en numerosas patologías. El análisis y modelado de la señal permite extraer de las imágenes médicas digitales la información de los biomarcadores mediante los procesos computacionales adecuados [Bonmati et al., 2012] reflejando la funcionalidad cerebral o los aspectos de los cambios patológicos “in vivo”. Por ejemplo mientras en SPECT, el biomarcador de imagen es la cantidad de fluido sanguíneo [Habert et al., 2011], en FDG-PET es el consumo de glucosa [Pupi et al., 2009], y ambos se ven decrementados en las

regiones donde se detecta primero la AD. En imagen de resonancia magnética estructural (MRI, del inglés “Magnetic Resonance Imaging”) la atrofia del lóbulo medio temporal (incluyendo el hipocampo y el cortex entorrinal) es el principal marcador para la detección precoz de la AD [Allan et al., 2010].

2.5 Diagnóstico de la AD: tests neuropsicológicos

Uno de los avances más importantes en el estudio contemporáneo de la AD ha sido el desarrollo y la amplia adopción del criterio de diagnóstico clínico. El libro *Diagnosics and Statistical Manual of Mental Disorders (DSM – IV)* [Association et al., 2000] explica que la AD conlleva una pérdida de las habilidades intelectuales de la suficiente severidad para interferir en las habilidades sociales u ocupacionales.

Un grado mayor de especificación fue realizado posteriormente mediante el test NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer’s Disease and Related Disorders Association) en 1984 [Dubois et al., 2007]. El criterio NINCDS-ADRDA especificó los factores de inclusión y exclusión abarcando tres niveles de confianza: probable, posible y definitivo. En 1997, se mejoraron los criterios neuropatológicos Khachaturian [Khachaturian, 2007] mediante los criterios *National Institute on Aging (NIA)-Reagan* [Sperling et al., 2011]. Estos esfuerzos han refinado las descripciones del diagnóstico clínico de la AD, facilitando la identificación de las correlaciones clínicas-patológicas. En las dos décadas pasadas, se ha realizado un progreso significativo para la estandarización de estos criterios de diagnóstico en términos de fiabilidad, sensibilidad y especificidad. En la actualidad, la AD se evalúa usando una amplia variedad de escalas validadas [Díaz and Haro, 2004], incluyendo medidas de:

1. Cognición: mediante el examen del Estado Mental Mínimo (del inglés “Mini Mental State Examination”, MMSE), o la Escala de Evaluación para la Enfermedad de Alzheimer (del inglés “Alzheimer’s Disease Assessment Scale-cognitive subscale”, ADAS-cog) o las Actividades de la vida diaria (del inglés “Activities of Daily Living”, ADL) o bien la Escala de Evaluación de Discapacidad para la demencia (“Disability Assessment for Dementia”, DAD).
2. Comportamiento: mediante inventario neuropsiquiátrico (del inglés “Neuropsychiatric Inventory”, NPI), o la patología de comportamiento en la escala de índice de la AD (del inglés “Behavioral Pathology in Alzheimer’s Disease Rating Scale”, BEHAVE-AD).

3. Funcionamiento global: mediante la entrevista del médico basada en la impresión de cambio (del inglés “Clinician’s Interview-Based Impression of Change”, CIBIC), Clasificación Clínica de las Demencias (del inglés “Clinical Dementia Rating Scale”, CDR).

El MMSE es probablemente el método más popular para visualizar el desajuste cognitivo mediante 30 ítems [Brugnolo et al., 2009] aportando resultados de fiabilidad entre el 80% y el 95%. Para medir el cambio cognitivo desde etapas tempranas a moderadas de la AD, la escala ADAS-cog ha sido la más utilizada [Drótos et al., 2012]. Se compone de 11 ítems cognitivos que incluyen la habilidad del lenguaje hablado, comprensión de lenguaje hablado, recordar instrucciones de prueba, dificultad en encontrar palabras en la conversación espontánea, seguir comandos, nombrar objetos y dedos, praxis constructivas e ideacional, orientación, tarea palabra-recuerdo, y tarea de reconocimiento de palabra. Se ha demostrado que el test de colinesterasa AchEIs tiene mayores beneficios en esta medida en etapas tempranas de la enfermedad, mientras que el más común para etapas moderadas a severas es la Batería de desajuste severo (SIB, del inglés “Severe Impairment Battery”).

Para encontrar especificaciones que regulen el tratamiento de los pacientes con AD, es necesario una evaluación clínica global con instrumentos tales como la CIBIC-Plus. Por otro lado, los síntomas neuropsiquiátricos son particularmente importantes ya que emergen durante los diferentes estadios de la enfermedad y su progreso suele estar en paralelo con la decadencia cognitiva. Se suelen evaluar con el NPI los síntomas de apatía, agitación, ansiedad, irritabilidad, desinhibición, alucinación, euforia entre otros [Feldman, 2007]. Finalmente, la discapacidad funcional se evalúa con numerosos tests como la escala de evaluación de discapacidad para demencia (DAD, del inglés “Disability Assessment for Dementia”), la escala de deterioro progresivo (PDS, (“Progressive Deterioration Scale”) o el estudio de las actividades de la vida diaria por parte de la cooperativa de la AD (“Alzheimer’s Disease Cooperative Study-Activities of Daily Living”, ADCS-ADL).

Algunos estudios han investigado la validez de estos criterios comparándolos con el diagnóstico de la AD postmortem. El promedio informó de sensibilidad, especificidad y exactitud general de diagnóstico que varió entre 70% y 90% [I. Goethals and Dierckx, 2002]. De esta forma, para garantizar el diagnóstico clínico, se necesitan tests adicionales para encontrar altos estándares. En 1996, la Academia Americana de Neurología recomendó SPECT como una técnica establecida para apoyar el diagnóstico clínico de la AD, basado en la observación de la perfusión decreciente en el corte de asociación del lóbulo parietal y las regiones temporales posteriores [I. Goethals and Dierckx, 2002].

La descripción de las técnicas de neuroimagen se lleva a cabo en el siguiente capítulo.

2.6 Tratamiento de la AD

El tratamiento esencial para la AD no ha sido establecido aún. Sin embargo, es posible retrasar la progresión de los síntomas reforzando de forma efectiva el tratamiento clínico y no-clínico que está basado en el diagnóstico temprano de la AD. La AD puede tratarse con terapias sintomáticas o que modifican la enfermedad como las neuroprotectoras o neurorestauradoras. Estas últimas resultan más eficaces al combinar los resultados clínicos y las medidas de biomarcadores [Cummings, 2009].

Los tratamientos actuales para pacientes con AD tienen como objetivo atender las opciones bioquímicas que estén asociadas con la enfermedad. Entre ellos se encuentra: la vacuna con $A\beta$, los antiagregantes de amiloide, antioxidantes como la vitamina E, factores neurotrópicos como Cerebrolysin (FPF-1070) [Guekht et al., 2011] etc. Los tratamientos sintomáticos actuales pueden mejorar el peso de la enfermedad con un impacto positivo en la cognición, función y el inventario neuropsiquiátrico. Este efecto positivo se ve incrementado cuanto antes se detecta la patología. Es por ello que a día de hoy, el mayor avance para combatir la AD es la detección temprana.

A nivel de investigación, los estudios estructurales y de neuroimagen son las únicas metodologías que permiten el estudio *in vivo* de la patología cerebral a un nivel macro y micromolecular. En AD, las regiones cerebrales características muestran un metabolismo de glucosa decrementado para FDG-PET y de flujo sanguíneo cerebral para SPECT, específicamente en los lóbulos temporoparietales, giro posterior cingulado y precuneiforme, así como en el corte frontal y la totalidad cerebral en pacientes afectados más severamente [Villemagne et al., 2005]. Estas anomalías son cambios mínimos en la etapa precoz de la AD y hacen del diagnóstico visual una tarea difícil incluso para médicos expertos lo cual ha potenciado las tecnologías de minería de datos para sistemas de diagnóstico asistido por ordenador (CAD).

2.7 Futuro de la AD

Con el incremento de la población perteneciente a la tercera edad, también ha aumentado la prevalencia de la AD; por lo que es más urgente tratar y diagnosticar correctamente la enfermedad por sus implicaciones sociales y económicas. De hecho, se ha progresado más en los últimos 15 años en la elucidación de la

causa y tratamiento de la AD que en el siglo pasado, lo que augura un paso más acelerado en la investigación [Gauthier, 2001].

Los avances en neuroimagen han provisto nuevas inclusiones en los cambios patológicos de la AD como por ejemplo el estudio de los diferentes biomarcadores para el diagnóstico temprano y la monitorización de la progresión de la enfermedad [Bonmati et al., 2012]. En lo que a genómica se refiere, las mutaciones en el cromosoma 21, 14 y 1 (APP, PSN1 y PSN2) han encabezado la investigación. Aparte de la identificación de la APOE, la búsqueda de otros genes determinantes no ha sido tan fructífera como se esperaba, aunque trabajos relacionados con el examen de polimorfismos nucleótidos se han provisto de herramientas útiles para la disección de los factores de riesgo de la AD [Reinikainen et al., 1989]. Este estudio de proteínas o proteómica está favoreciendo el desarrollo de los biomarcadores de la AD así como los objetivos terapéuticos.

La línea de investigación en un futuro consistiría en retraer los determinantes genómicos de la expresión de la proteína anormal y considerar la terapia farmacogénica en un estado temprano de la enfermedad antes de que los cambios neuropatológicos ocurran. Los intentos más cercanos para estudiar el progreso de la AD se han concentrado alrededor del CSF y de las técnicas de neuroimagen (descritas en el capítulo 3 de este trabajo). Tres de los biomarcadores más prometedores que se encuentran en el CSF (P-tau, T-tau y β -amiloide) tienen una alta sensibilidad para diferenciar AD temprana de vejez normal, depresión, alcoholismo, Parkinson; pero menor especificidad frente a otras demencias como la degeneración frontotemporal y la demencia de los cuerpos de Lewy [Marui et al., 2002]. El hecho de combinar un conjunto de biomarcadores CSF puede mejorar la exactitud de diagnóstico, por ejemplo combinando CSF β -amiloide y tau total se ha llegado a alcanzar una especificidad del 85% [Feldman, 2007]. En conclusión, el futuro de la AD durante los próximos 10 años, se dirige en avanzar tanto en detección temprana como en terapia y comprensión de la AD [Hampel et al., 2011].

Parte II

Estado del arte

Capítulo 3

Técnicas de obtención de imágenes médicas

Este capítulo expone las técnicas de neuroimagen utilizadas en la actualidad en medicina nuclear, disciplina que facilita el diagnóstico y tratamiento de determinadas enfermedades (por ejemplo la AD) mediante la utilización de medicamentos radiofármacos. Según la visión del cerebro y la evidencia del cambio a lo largo del tiempo, las técnicas de neuroimagen se clasifican en estructural y funcional, de las que se describen sus fundamentos, analogías y diferencias. En lo que concierne a la prevención temprana de la AD, la técnica de imagen funcional resulta ventajosa para evaluar los procesos biológicos “in vivo”. Además dado que la experimentación de este trabajo se lleva a cabo con las categorías de imágenes funcionales PET y SPECT, se dedica especial atención a sus fundamentos, a las cámaras utilizadas en su adquisición y a las técnicas de preprocesado tales como la reconstrucción y la normalización (espacial y en intensidad).

3.1 Neuroimagen estructural y funcional

Las técnicas de neuroimagen permiten ver imágenes “in vivo” del sistema nervioso central en general y del cerebro en particular aportando información complementaria de gran importancia para el diagnóstico y tratamiento de pacientes con trastornos del sistema nervioso [Martí-Climent et al., 2010]. Las técnicas de neuroimagen pueden clasificarse en estructural o funcional, según la rapidez de la evidencia del cambio a lo largo del tiempo y la visión del cerebro que ofrecen [Villemagne et al., 2005].

La imagen estructural ofrece una visión de la estructura cerebral para permitir el diagnóstico de enfermedades a gran escala. Es decir, los cambios estructurales no son evidentes hasta tarde durante el curso de la enfermedad, lo que ha motivado el desarrollo y refinamiento de técnicas más sofisticadas, como la imagen volumétrica en serie y la substracción de la compresión de vóxel, enfatizando una aproximación cuantitativa capaz de revelar cambios sutiles a lo largo del tiempo [Symms et al., 2004].

Por otro lado, las técnicas de neuroimagen funcional tales como SPECT y PET son muy valiosas para el diagnóstico diferencial de la AD, así como para la valoración del pronóstico. Con la llegada de estrategias terapéuticas que tienen como objetivo reducir la carga del $A\beta$ en el cerebro, hay un interés creciente en los radioligandos de SPECT y PET que permiten la evaluación de la carga del $A\beta$ “in vivo” en conjunción con otros biomarcadores $A\beta$ en el plasma y CSF [Villemagne et al., 2005]. La figura 3.1 muestra la variación de anormalidades conforme avanza la AD para los tipos de imágenes funcional FDG-PET y estructural MRI. Para FDG-PET, se espera que las anormalidades aparezcan en el siguiente orden: precuneus/cingulado posterior primero, lateral temporal y lóbulo frontal mucho más tarde. De forma similar, en MRI (estructural), las anormalidades se esperan en el orden: medio temporal, temporal lateral y lóbulo frontal. Se observa que los cambios anormales relativos a la AD en imagen funcional se detectan antes que para MRI estructural.

Las técnicas de neuroimagen estructural [Symms et al., 2004] tales como la CT y la MRI, se utilizan rutinariamente en la evaluación clínica de los pacientes con AD. La CT se utiliza principalmente para excluir otras causas tratables de demencia. Los estudios que informan de los cambios relacionados con la AD dieron a conocer que todos los grupos de demencia tendían a mostrar en CT la atrofia de los lóbulos temporales medios (MTL, del inglés “Medial Temporal Lobe”) pero esta atrofia fue incapaz de discriminar entre los diferentes tipos de demencia. Por otro lado, MRI [Vemuri et al., 2008] se ha utilizado para examinar la atrofia en la corteza entorrinal y en cortes temporales en pacientes con AD temprano, donde la severidad de la pérdida del

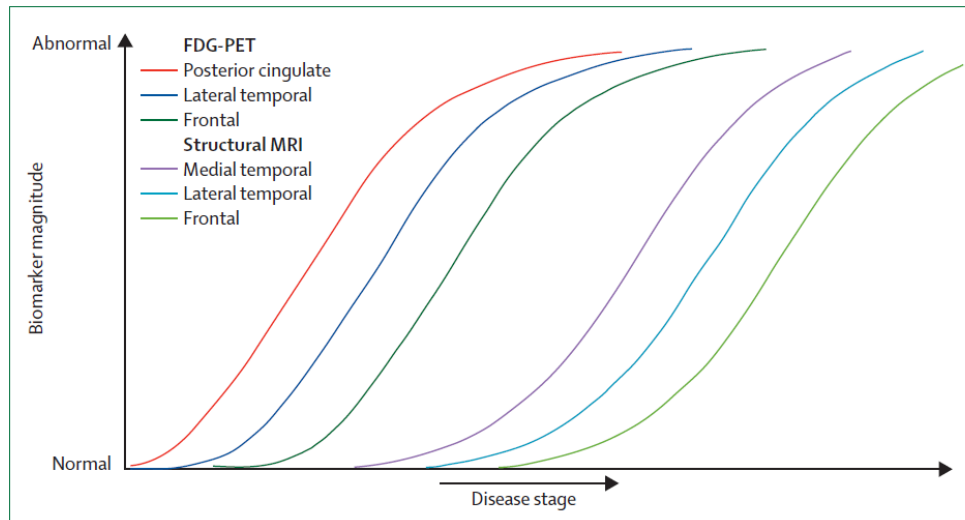


Figura 3.1: Información anatómica de la imagen. Comparación de la magnitud de los biomarcadores funcionales frente a los estructurales teniendo en cuenta el avance de la enfermedad de Alzheimer. Obtenida de [Wechalekar and Cook, 2005]

volumen está relacionada con la severidad de la enfermedad. Estas regiones se redujeron tanto en MCI como en AD en comparación con los normales. Ambas condiciones se relacionaron estrechamente con la pérdida de la materia gris y con el aumento ventricular. Otros estudios sugirieron que las medidas de los cambios en el hipocampo fueron más prácticos y útiles. La pérdida de volumen en el hipocampo y en el parahipocampo fue significativamente mayor en pacientes con AD que en pacientes con DLB (Demencia con cuerpos de Lewy) [Marui et al., 2002] que es consistente con la preservación de la memoria observada en DLB.

La imagen funcional se utiliza para diagnosticar tumores más pequeños y enfermedades a una escala menor. La imagen funcional también permite visualizar la relación entre la actividad en ciertas áreas cerebrales y las funciones mentales específicas. La imagen funcional a menudo se utiliza en investigación científica neurológica y cognitiva. Las modalidades de imagen funcional más sensibles tales como la imagen funcional de resonancia magnética (fMRI) [Courtney et al., 1997], perfusión MR, espectroscopía MR (MRS) [Villemagne et al., 2005], imagen de difusión pesada MR, PET y SPECT tienen la capacidad de identificar cambios patopsicológicos en el cerebro, antes de que los

cambios estructurales estén presentes por lo tanto presentando un mayor potencial para un diagnóstico precoz exacto, monitorizando la progresión de la enfermedad y una mejor consecución del tratamiento. Mientras que fMRI detecta cambios regionales en la concentración de la deoxihemoglobina reflejando la actividad focal cerebral en respuesta a una variedad de estímulos y tareas de memoria, la perfusión MR muestra sensibilidad y especificidad en torno al 90% para pacientes MCI [Courtney et al., 1997]. Por otro lado, hay estudios MRS que muestran decrementos en NAA (N-acetil aspartato) en pacientes con AD en los cortes temporales y parietales, demostrando que la reducción de NAA y el aumento de la severidad de la enfermedad por criterios neuropatológicos están relacionados.

Los pacientes con AD muestran una difusión media significativamente mayor en el hipocampo [Bouras et al., 1994], cíngulo, materia blanca temporal y parietal que los controles usando imagen MR pesada de difusión, mientras que la imagen MR de tensor de difusión ha mostrado la reducción difusa en la integridad de la materia blanca para pacientes con AD.

Tanto las técnicas SPECT como PET son de imagen funcional molecular y utilizan trazadores radioetiquetados para evaluar los procesos biológicos *in vivo*. Estas técnicas pueden permitir la detección de los procesos de la enfermedad en fases asintomáticas cuando no existe evidencia de los cambios anatómicos en CT y MRI [Villemagne et al., 2005] como se observa en la figura 3.2. La figura 3.2 muestra la comparación entre las imágenes funcionales (parte superior) y estructurales (parte inferior) siguiendo los diferentes estadios de la AD: a) individuo normal sin evidencia de $A\beta$ ni de atrofia MRI, b) individuo cognitivamente normal sin evidencia de atrofia en MRI pero con depósito significativo de $A\beta$, c) individuo con demencia y diagnóstico clínico AD mediante un estudio positivo de imagen PET amiloide, y atrofia neurodegenerativa en MRI. Esta figura apoya el principio de que sólo la presencia del depósito de $A\beta$, incluso en cantidades importantes, no es suficiente para detectar demencia (como se muestra en 3.2(b)), y que las anomalías en biomarcadores del depósito de *Abeta* preceden los síntomas clínico/cognitivos.

3.2 Medicina nuclear

La Medicina Nuclear se define como la rama de la medicina que emplea los isótopos radiactivos, las radiaciones nucleares, las variaciones electromagnéticas de los componentes del núcleo atómico y técnicas biofísicas afines, para la prevención, diagnóstico, terapéutica e investigación médica. Sus principales campos de acción son el diagnóstico por imagen y el tratamiento de deter-

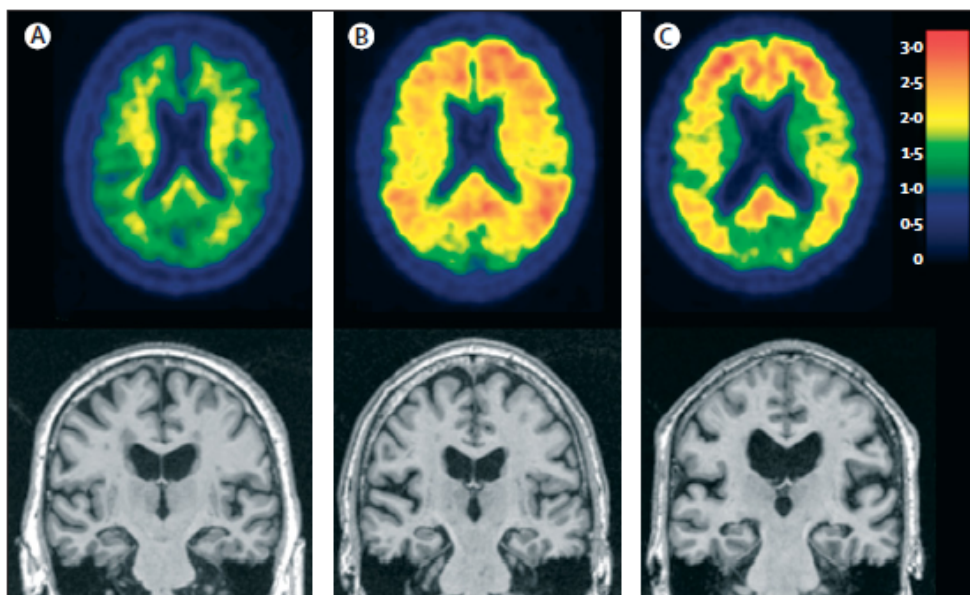


Figura 3.2: Etapas de los biomarcadores en la AD. a) Un individuo cognitivamente normal sin evidencia de $A\beta$ en imágenes de amiloide PET con PiB y sin evidencia de atrofia en MRI. b) Un individuo cognitivamente normal sin evidencia de atrofia neurodegenerativa en MRI pero con depósito significativo $A\beta$ en las imágenes PET. c) Un individuo con demencia y un diagnóstico clínico de la AD, un estudio positivo de imagen de amiloide PET y una atrofia neurodegenerativa en MRI. Obtenida de [Jack et al., 2010].

minadas enfermedades mediante el uso de medicamentos radiofármacos. Las aplicaciones clínicas de los radiofármacos abarcan prácticamente a todas las especialidades médicas ¹.

También denominados radioisótopos, radionúclidos o radionucleidos; los isótopos radiactivos reúnen dos propiedades que justifican su aplicación clínica [Díaz and Haro, 2004]:

1. Los mecanismos de captación y metabolización en el organismo no se modifican ya que sus propiedades químicas son las mismas que las del elemento estable, no radiactivo.
2. Al desintegrarse, los radionúclidos emiten radiación, la cual permite conocer la localización y distribución corporal del isótopo.

3.3 Neuroimagen funcional

En esta sección describimos las técnicas de neuroimagen funcional utilizadas en este trabajo (SPECT y PET) que han producido los mayores avances en la AD durante las dos pasadas décadas.

3.3.1 SPECT cerebral de perfusión

Esta exploración estudia la distribución regional del flujo sanguíneo en el cerebro. La distribución del flujo no es homogénea, la sustancia gris recibe cuatro veces más sangre que la sustancia blanca. La SPECT [Domínguez-Gadea, 2008] de perfusión es, indirectamente, una fuente de información sobre la actividad funcional y metabólica de distintas áreas cerebrales ya que el aporte se distribuye en función de la actividad metabólica de cada área cerebral, ajustando la oferta de sangre a la demanda celular.

Respecto a los radiofármacos utilizados, el primer requisito que un trazador debe cumplir es su lipofilia. Las sustancias liposolubles se difunden pasivamente a través de la membrana de las células endoteliales y logran alcanzar el cerebro, dado que son capaces de atravesar la barrera hematoencefálica (BHE) [Díaz and Haro, 2004].

La hexametilpropilenoaminoxima (HMPAO) fue propuesta en 1985 por Nowotnik como molécula grasa que consigue cruzar la BHE intacta por difusión pasiva. Su permanencia en el cerebro se debe a su rápida metabolización por las neuronas, sufriendo un proceso de oxidación que provoca la pérdida de su lipofilia inicial y queda atrapada en ellas. El HMPAO fue el primer trazador

¹Definición adoptada por la Comisión Nacional de la Especialidad de Medicina Nuclear

Comparación entre los principales trazadores de SPECT cerebral.		
	HMPAO	Tc-99m
Vida media (h)	72	24
Eficiencia de extracción al primer paso (%)	80	70
Redistribución cerebral	No	No
Actividad en sangre en 1 h (%)	12	4.9
Estabilidad tras el marcaje (h)	0.5	mayor que 6
Contraste de imagen	Menor	Mayor
Dosis efectiva cuerpo entero (mSr)	5-6.2	4.3

Tabla 3.1: Comparación entre los principales trazadores de SPECT cerebral. Obtenida de [Díaz and Haro, 2004]

tecnecio disponible comercialmente [Slosman et al., 2001]. Además del HMPAO, también se utilizan otros radiofármacos como el Tc99m-ECD (tecnecio 99m etilencisteinatodímero, Neurolite) que carece de los problemas de inestabilidad “in vitro” del HMPAO, ya que permanece estable durante 6 h (su emisión gamma de 140 keV es la ideal para la detección en la gammacámara) desde el marcaje a temperatura ambiente lo que permite realizar cómodamente el estudio sin prolongar en exceso la emisión radiactiva, y además consigue imágenes de mayor calidad al lograr un mayor contraste entre la sustancia gris y sustancia blanca (contraste interno) o entre el tejido cerebral y los tejidos vecinos (contraste externo, el mayor aclaramiento sanguíneo reduce la actividad de fondo) [Díaz and Haro, 2004]. También su disponibilidad es cómoda y económica a partir del generador $^{99m}\text{Tc}/^{99}\text{Mo}$. Su administración se realiza por vía intravenosa, lo que permite que en sólo 20 a 30 minutos se concentre en el tiroides en cantidad suficiente para que se pueda obtener la imagen. Todos estos detalles se encuentran resumidos en la tabla 3.1, donde encontramos una comparativa entre Tc99m-ECD y el HMPAO.

En la figura 3.3, se representan varios cortes coronales de la base de datos funcional SPECT utilizada en esta tesis explicada en la sección 8.1. Se observa cómo la distribución regional sanguínea de un sujeto normal (primera fila) varía con respecto a un paciente AD (paciente 2) donde aparece reducida sobre todo en las zonas afectadas por el AD (más detalles en el capítulo 8).

3.3.2 PET

La PET es una técnica de diagnóstico dentro del campo de medicina nuclear que se basa en la obtención de imágenes de la distribución del organismo de

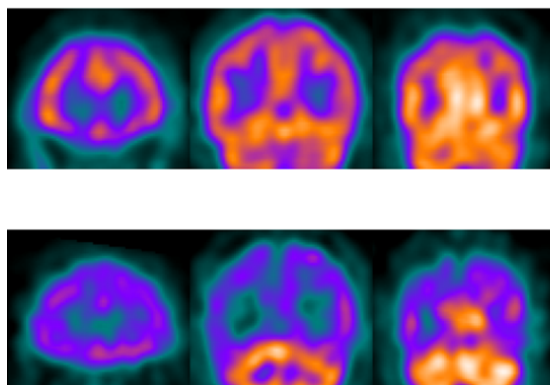


Figura 3.3: Patrones de perfusión cerebral regional con SPECT en sujeto sano (primera fila) y paciente AD (segunda fila).

determinadas moléculas (radiofármacos) marcadas con isótopos emisores de positrones [Sánchez, 2001]. Entre las ventajas de esta técnica deben destacarse que no es invasiva, no tiene efectos secundarios, ni complicaciones, y que es posible realizar un estudio de interés, bien de una parte o de todo el cuerpo en una sola exploración y en un tiempo aproximado de 40 minutos.

El instituto americano de salud ha fundado la iniciativa de neuroimagen para la AD (*Alzheimer's Disease Neuroimaging Initiative*, ADNI) [Weiner et al., 2012] con el objetivo de establecer estándares para la adquisición de imágenes, así como para el desarrollo un repositorio de datos (escáneres de tipo MRI o PET que comentamos en esta sección e incluimos en la experimentación de este trabajo), y el análisis de las mejores aproximaciones de imagen para pruebas clínicas.

En la tabla 3.2 se presenta una comparación entre SPECT y PET. Entre las ventajas de PET se encuentra que consigue una buena resolución espacial y no precisa de colimadores de plomo para eliminar la radiación Compton. Otra de las grandes ventajas de PET es que los radionúclidos emisores de positrones son isótopos de elementos biológicos constituyentes del cuerpo humano, lo que permite estudiar cualquier proceso fisiológico. Casi todos están producidos por ciclotrón y los más frecuentes son: el carbono (^{11}C), el nitrógeno (^{13}N), el oxígeno (^{15}O) y especialmente el flúor (^{18}F), siendo la fluorodeoxiglucosa ^{18}F (^{18}F)-FDG el radiofármaco más utilizado en estudios PET [Díaz and Haro, 2004]. Éste una vez inyectado, se distribuye por todas las células del

Comparación entre SPECT y PET		
	SPECT	PET
Radionúclido emisor	De fotones	De positrones
Vida media del radionúclido	Horas–días	Minutos
Estudio funcional	Sí	Sí
Reconstrucción tomografica y funcional	Sí	Sí
Marcaje	Tradicional	Automatizado
Resolución espacial (mm)	7–6	5–6
Sensibilidad	Alta	Muy alta
Necesidad de colimador y ventana	Sí	No
Disponibilidad/coste	Asequible	Poco asequible

Tabla 3.2: Comparación entre PET y SPECT (obtenida con HMPAO). Extraída de [Díaz and Haro, 2004]

organismo, concentrándose donde es mayor el metabolismo celular. Este hecho permite estudiar cualquier órgano o tejido con aumento del metabolismo de glucosa.

El principal inconveniente de PET es su escasa disponibilidad dado su elevado coste. Este aspecto está condicionado por el corto período de semidesintegración de los radionúclidos emisores de positrones que obliga a contar con un ciclotrón *in situ*, obteniéndolos por tanto en el mismo centro hospitalario. Además PET no detecta exactamente la localización del positrón ni del trazador, sino del par de fotones de aniquilación que se genera provocando una disminución en la resolución.

Una nueva era se ha abierto con el desarrollo de radioligandos que pueden ser administrados de forma segura a los humanos y que pueden etiquetar proteínas de agregado incluyendo β -amiloide y tau. Se dispone en la actualidad de dos trazadores de este tipo: el compuesto Pittsburgh-B (PIB con etiqueta de carbono 11) [Illán et al., 2010] que etiqueta el β -amiloide, y el [18F]FDDNP (flúor-18 etiquetados FDDNP), que etiqueta tanto el β -amiloide como el tau [Shin et al., 2010]. Ambos se han probado en humanos y se ha demostrado que la característica de retención en sujetos con AD es consistente con la patología. El potencial de estos compuestos recae en su habilidad para cuantificar la AD en vivo por primera vez. La combinación de estos compuestos con otros biomarcadores bioquímicos y genéticos es un aditivo potente para la batería actual de biomarcadores. En la actualidad, ligandos con el trazador β amiloide y tau se están desarrollando para SPECT aunque esto se ha limitado para estudios *in vitro* hasta el momento.

3.3.3 Tomografía computarizada

La tomografía computarizada (CT, del inglés “Computed Tomography”) ha sido usada extensamente para la investigación de la AD [Pfefferbaum et al., 1990]. Involucra la proyección de datos provenientes de múltiples direcciones y la reconstrucción tomográfica, por ejemplo, las emisiones radioactivas de los trazadores descritos en la sección anterior.

Para la obtención de distintas proyecciones radiográficas desde distintos ángulos en la CT, el tubo de rayos X y los detectores no se mantienen fijos, sino que rotan síncronamente y en posición enfrentada alrededor del paciente, montados sobre un raíl anular llamado *gantry*. Una vez adquirido un número suficiente de proyecciones, se utiliza el método informático clásico de retro-proyección filtrada, que calcula el coeficiente de atenuación lineal promedio de cada vóxel. El equipo construye la imagen final asignando a cada píxel un valor de intensidad en función del índice de atenuación promedio del vóxel que representa y unas coordenadas espaciales, que determinan su posición en la imagen. El detector (*gantry*) rota alrededor del objeto de interés y permite poder observar el patrón de emisión en el campo de visión desde muchos ángulos [Bruyant, 2002] como se muestra en la figura 3.4. La formación de imágenes tomográficas, es decir, la que corresponde a un plano o corte, a partir de la radiación emitida por el trazador localizado en el interior del organismo, recibe la denominación genérica de CT por emisión. Esta presenta dos variedades según el tipo de radionúclido utilizado: SPECT y PET.

3.3.4 Cámaras para la obtención de imágenes funcionales

Cámaras SPECT

SPECT se basa en la detección del único fotón procedente de la desintegración del radionúclido. Existen varios tipos de cámaras SPECT, por ejemplo, la cámara Anger [Graham et al., 2002], gira 360 grados alrededor del paciente mientras adquiere imágenes planares desde distintos ángulos. La cámara para la obtención de SPECT en esta tesis es la Picker Prism 3000 ² (del Hospital Universitario Virgen de las Nieves, Granada) y posee tres cabezales que giran 120 grados (por cabezal) alrededor del paciente para la adquisición de las imágenes planares desde distintos ángulos. Habitualmente se adquieren imágenes cada 3 o 6 grados [Christensen and Uzzel, 2000], en nuestro caso, se toman 180 proyecciones con una resolución angular de 2 grados. Cuanto mayor es el número de proyecciones y el tiempo de adquisición dedicado a cada

²<http://www.eclipsesys.com/PsychiatrySystems/SPECTPicker3000XPApplications/>

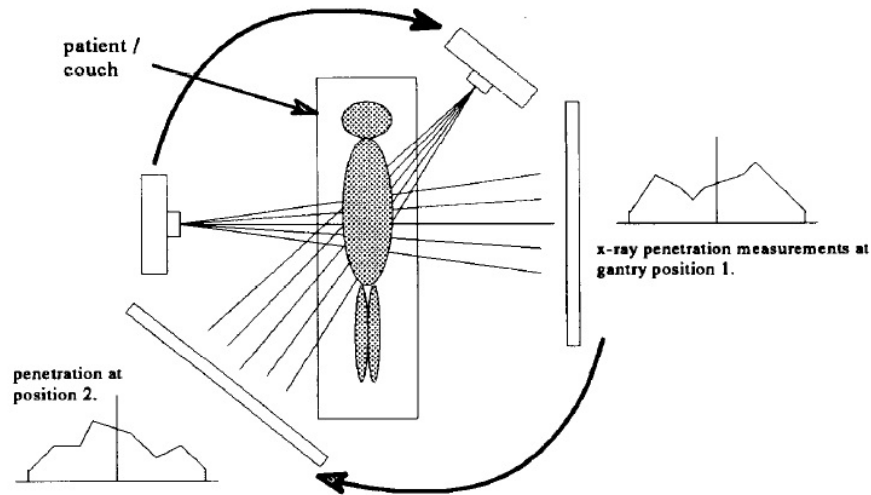


Figura 3.4: Escáner CT mostrando un detector gantry rotativo para la adquisición de medidas en diferentes posiciones. Obtenida de [Wechalekar and Cook, 2005].

una de ellas, mayor será la resolución de la imagen, a expensas de invertir más tiempo en la adquisición.

Durante la adquisición, el paciente se posiciona cómodamente tumbado con la cabeza inmobilizada en el cabezal de la gammacámara. El detector se posiciona tan cerca de la cabeza del paciente como sea posible, preferiblemente con un radio de rotación de 14 cm o menos desde la superficie del colimador al centro del cerebro del paciente. A los pacientes se les inyecta el dímero de etil-cisteína marcado con tecnecio-99m y el escáner SPECT obtiene las proyecciones basadas en número de cuentas en la emisión de fotones. Finalmente, las imágenes de las secciones cruzadas cerebrales de los datos de proyección se reconstruyen usando el algoritmo FBP en combinación con un filtro de Butterworth para eliminar el ruido.

Cámaras PET

El gran interés mostrado hacia la instrumentación PET se debe fundamentalmente al hecho de que con esta técnica se pueden lograr imágenes funcionales reales de calidad. Además, con la introducción de cristales de centelleo nuevos y de sistemas PET/CT, se puede aumentar el rendimiento del escáner [Tarrantola et al., 2003]. La principal diferencia con SPECT radica en que esta uti-

liza radionucleidos emisores de fotón simple (que emiten rayos gamma normales como el tecnecio-99m), mientras que en PET se emplean radionucleidos emisores de positrones (lo cual implica detección simultánea de 2 rayos gamma por parte de 2 detectores opuestos).

Antes de la adquisición, el paciente debe ayunar al menos 4 horas en las que se sitúa tumbado relajadamente con la cabeza situada a la entrada de la cámara PET. Tras la inyección del radionúcleo, el sujeto debe permanecer tumbado cómodamente con los ojos abiertos en la habitación durante 20 minutos para la incorporación del radiofármaco ([18F]-FDG o C-PIB en nuestro caso) en el cerebro. Una vez que el radiofármaco se encuentra en el interior del paciente, los positrones emitidos (β^+), tras un recorrido pequeño (2–3 mm, son atraídos fuertemente por los electrones (e^-) del medio, que los aniquilan generando una energía en forma de par de fotones de 511 eV de la misma dirección y sentidos opuestos. La PET se basa por tanto en la detección de dos fotones de aniquilación mediante un cristal de centelleo.

Para ello dispone de un *gantry* que contiene varios anillos (entre 4 y 16) de fotodetectores responsables de detectar los dos fotones de aniquilación [Díaz and Haro, 2004], en el caso de la iniciativa ADNI se trata de una cámara Siemens *HR/HR + ECAT/ACCEL*. Desde su emisión hasta alcanzar el fotodetector, uno o los dos fotones puede sufrir el efecto Compton [Yoon et al., 2011] afectando a la calidad de la imagen. Cada detector está conectado por un circuito de coincidencia con varios de los situados enfrente, así el sistema puede valorar si los fotones han llegado en la misma dirección, con la misma energía (511 keV) y en coincidencia, o en otras palabras en un intervalo de tiempo inferior a 1 nanosegundo.

3.4 Técnicas de preprocesado de imágenes

Las imágenes raw obtenidas directamente de la gammacámara necesitan ser procesadas, en primer lugar, mediante un procedimiento de reconstrucción (para recuperar imágenes cerebrales en 3D a partir de los datos de proyección) y en segundo, con procedimiento de normalización (para que exista la misma correspondencia anatómica y de intensidad en imágenes de diferentes pacientes). La figura 3.5(a) representa una imagen raw (del total de 180 imágenes de proyecciones) obtenida de la gammacámara SPECT Picker Prism 3000 de Hospital Universitario Virgen de las Nieves (Granada). A continuación describimos algunos algoritmos utilizados para la reconstrucción.

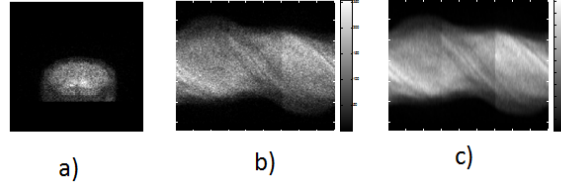


Figura 3.5: a) Una imagen raw SPECT para un control (de un total de 180 proyecciones obtenidas cada 2 grados), b) sinograma SPECT, c) sinograma SPECT filtrado.

3.4.1 Reconstrucción por retroproyección filtrada

Una imagen de la sección cruzada de un objeto puede ser recuperada o reconstruída a partir de los datos de proyección. En condiciones ideales, las proyecciones son un conjunto de medidas de los valores integrados de algún parámetro del objeto [Bruyant, 2002], [Shepp and Vardi, 1982]. Si el objeto está representado por una función bidimensional $f(x, y)$ que representa el número de fotones emitidos en cualquier punto (x, y) del corte transversal en el campo de visión, y cada integral de línea por los parámetros (θ, t) , la integral de línea P_θ se define como:

$$P_\theta(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x \cdot \cos(\theta) + y \cdot \sin(\theta) - t) dx dy \quad (3.1)$$

donde la función $P_\theta(t)$ se conoce como la transformada de Radon (también llamada sinograma) [yan Qu et al., 2012] de la función $f(x, y)$. La figura 3.5(b) muestra el sinograma de la imagen SPECT raw (representada en 3.5(a)). El sinograma equivale a 180 proyecciones (1 cada 2 grados) obtenidas de la cámara SPECT). En 3.5(c), se representa el sinograma con la corrección de movimiento (efecto indeseado de la gammacámara). Una proyección está formada por la combinación de un conjunto de integrales de línea. La clave de la imagen tomográfica es el teorema de cortes de Fourier que relaciona la proyección del dato medido con la transformada de Fourier bidimensional de la sección del objeto. La transformada de una proyección paralela $P_\theta(t)$ de un corte de imagen $f(x, y)$ tomado en un ángulo θ se define con las ecuaciones 3.6 y 3.3:

$$S_\theta(\omega) = \int_{-\infty}^{+\infty} P_\theta(t) \cdot \exp(-j \cdot 2\pi \cdot \omega \cdot t) dt \quad (3.2)$$

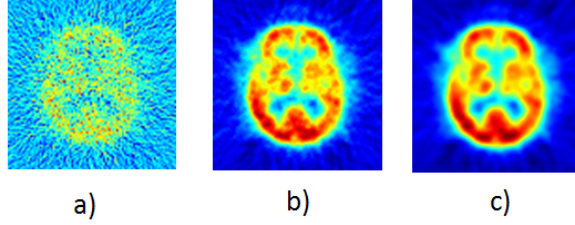


Figura 3.6: Preprocesado de un corte coronal de la base de datos SPECT en sus fases de: a)prefiltrado, b)reconstrucción FBP, c)postfiltrado.

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \cdot \exp(-j \cdot 2\pi \cdot \omega \cdot t) dx \cdot dy \quad (3.3)$$

Tomando proyecciones de la función de un objeto en ángulos $\theta_1, \theta_2, \dots, \theta_k$ y su transformada de Fourier, los valores de $F(u, v)$ pueden ser determinados en líneas radiales, por lo que $S_\theta(\omega)$ se define como:

$$S_\theta(\omega) = F(u = \omega \cdot \cos(\theta), v = \omega \cdot \sin(\theta)) \quad (3.4)$$

El algoritmo de retroproyección filtrada (FBP) [Bruyant, 2002] puede derivarse fácilmente del teorema de Fourier por cortes. Una imagen de la sección cruzada $f(x, y)$ de un objeto puede ser recuperada por:

$$f(x, y) = \int_0^\pi Q_\theta(x \cdot \cos(\theta) + y \cdot \sin(\theta)) d\theta \quad (3.5)$$

$$Q_\theta(t) = \int_{-\infty}^{+\infty} S_\theta(\omega) \cdot |\omega| \cdot \exp(j \cdot 2\pi \cdot \omega \cdot t) d\omega \quad (3.6)$$

El algoritmo FBP consta de dos fases [Bruyant, 2002], [Chornoboy et al., 1990]: el filtrado, que se puede visualizar como un simple pesado de cada proyección en el dominio de la frecuencia y la parte de retroproyección. La figura 3.6(a) muestra el prefiltrado de un corte coronal SPECT, mientras que la figura 3.6(b) representa la FBP y la 3.6(c) el posfiltrado. FBP ha sido utilizado para la reconstrucción de las imágenes utilizadas en este trabajo (descritas en el capítulo 8) y la mayor ventaja frente a MLEM (del inglés “maximum likelihood estimation maximization”) y OSEM (del inglés “ordered subset estimation maximization”) es su rapidez.

3.4.2 MLEM

El algoritmo MLEM para la reconstrucción de imágenes es iterativo y de los que mejor modelan los fotones emitidos por una fuente radiactiva logrando la mejor calidad de imagen (cuando se compara con FBP [Gaitanis et al., 2010]) aunque su convergencia es lenta.

Una densidad de emisión desconocida $\lambda = \lambda(x, y, z)$ debe ser reconstruída desde el número de cuentas $n^*(d)$ en cada uno de las D unidades de detectores. Dentro del modelo, existe un algoritmo para determinar la estimación $\hat{\lambda}$ de λ que maximiza la probabilidad $p(n^*|\lambda)$ de observar la cuenta del detector actual n^* sobre todas las posibles densidades λ [Shepp and Vardi, 1982]. Las variables independientes de Poisson $n(b)$ con densidades de emisión $\lambda(b)$, $b = 1, \dots, B$ representan el número de emisiones no-observadas en cada uno de los B píxeles en que se particiona un objeto que contiene un emisor. Supongamos que cada emisor en la caja B se detecta en una unidad de detección d con probabilidad $p(b, d) = 1$, $d = 1, \dots, D$ siendo $p(b, d)$ una matriz de transición de un paso que se asume conocida. Si observamos el número total $n^* = n^*(d)$ de emisiones en cada unidad de detector d y queremos estimar la $\lambda = \lambda(b)$, $b = 1, \dots, B$. Para cada λ , los datos observados n^* tienen una probabilidad $p(n^*|\lambda)$. El algoritmo EM de la estadística matemática comienza con una estimación inicial λ^0 y da el siguiente procedimiento iterativo simple para obtener una nueva estimación $\hat{\lambda}^{nueva}$ a partir de una estimación anterior $\hat{\lambda}^{antigua}$, para obtener:

$$\hat{\lambda}^{nuevo}(b) = \hat{\lambda}^{antiguo}(b) \cdot \sum_{d=1}^D \frac{n^*(d) \cdot p(b, d)}{\sum_{b=1}^B \hat{\lambda}^{antiguo}(b) \cdot p(b, d)} \quad (3.7)$$

3.4.3 OSEM

La reconstrucción de imagen llevada a cabo por OSEM agrupa las proyecciones de los datos en subconjuntos ordenados [Reilhac et al., 2008], [Alzimami et al., 2011], [Erlandsson et al., 2011]. El algoritmo OSEM (al iterar sobre conjuntos de datos) proporciona una magnitud de aceleración frente a MLEM manteniendo la calidad de reconstrucción. Aunque la calidad de la reconstrucción es buena, la aplicación de EM provoca un aumento de la carga computacional y una convergencia lenta, incluso con técnicas de aceleración estándar. Por eso se está optando por procesar los datos con algoritmos de subconjuntos ordenados (OS) que procesan los datos en bloques dentro de cada iteración y muestra que este procedimiento acelera la convergencia por un factor proporcional al número de subredes.

Definimos una iteración en OSEM como un único paso a través de todos

los subconjuntos especificados. Las iteraciones más lejanas se llevan a cabo pasando a través de los mismos subconjuntos ordenados, usando como punto de comienzo de la reconstrucción provista por la iteración previa. Con subconjuntos mutuamente exclusivos (y exhaustivos), cada iteración OSEM tendrá un tiempo de cómputo similar a una iteración EM estándar.

El algoritmo es de la siguiente forma [Hudson et al., 1994]. Se define la imagen $x = |x_j; j = 1, \dots, J|$. Se especifica un píxel individual y los elementos de detección (o rayos) por j y t respectivamente. Los pesos a_{tj} representan la probabilidad de que la emisión del píxel j se grabe en t . Luego las cuentas del detector se distribuyen por Poisson con valores de esperanza $\mu = E_y = Ax$, donde A es la matriz de proyección. Siendo \hat{x}^0 una imagen inicial y \hat{x}^m la estimación de x tras la iteración m . S_1, S_2, \dots, S_n denotan los subconjuntos para el detector en el orden seleccionado. Pasos del algoritmo:

1. $m=0, \hat{x}^m$.
2. Repetir hasta la convergencia de \hat{x}^m .
 - $x^1 = \hat{x}^m$
 - Para las subredes $i = 1, \dots, m$:
 - Proyección: calcular los valores esperados para las cuentas acumuladas como:

$$\mu_t^i = \sum_{j=1}^J a_{tj} \cdot x_j^i \quad (3.8)$$

para $t \in S_i$

- Retroproyección. Calcular:

$$x_j^{i+1} = x_j^i \cdot \sum_{t \in S_i} \frac{y_t \cdot a_{tj}}{\mu_t^i} / \sum_{i \in S_i} a_{tj} \quad (3.9)$$

- $\hat{x}^m = x^{n+1}$

3.4.4 Normalización

Para poder realizar un análisis correcto de las imágenes funcionales, necesitamos que la misma posición en imágenes diferentes se corresponda con la misma localización anatómica. Esto posibilita la comparación de valores de vóxel en una cierta posición para diferentes imágenes y para su posterior tratamiento [Stoeckel, 2003].

1. En primer lugar, el sujeto no está siempre posicionado en la misma posición en el “frame” de referencia que define dónde se posiciona el cerebro en la imagen. Este problema se extrapola tanto para imágenes del mismo sujeto como para imágenes de distintos sujetos.
2. En segundo lugar, los rasgos anatómicos no tienen la misma forma y tamaño entre sujetos diferentes (por ejemplo, el tamaño y forma del cráneo puede variar bastante entre sujetos). Este problema ocurre en general sólo entre imágenes de sujetos diferentes porque en la mayoría de los casos la anatomía no cambia significativamente entre adquisiciones distintas del mismo sujeto.

En nuestro caso, el objetivo de la normalización espacial es hacer que las posiciones anatómicas coincidan entre imágenes distintas. A continuación, se establecen los principios fundamentales de la normalización espacial en la que se inciden en los métodos de registro de imágenes y regularización desarrollados para estimar las transformaciones entre imágenes para que las posiciones anatómicas coincidan entre diferentes imágenes.

Posterior a la normalización espacial, es necesaria una normalización en intensidad dado que el nivel medio de metabolismo o perfusión cerebral varía enormemente entre distintos sujetos y, en menor medida, en un mismo sujeto con el tiempo. Por ello es necesario ajustar los datos de intensidad para efectuar análisis cuantitativos tanto longitudinales como transversales [Gispert et al., 2003].

3.4.5 Normalización espacial

El problema de la normalización espacial consiste en mapear las imágenes cerebrales de sujetos en un espacio estándar para poder realizar comparativas fiables entre ellos [Gispert et al., 2003]. La naturaleza de esta comparación se determina por la modalidad de las imágenes y lo que se pide de los datos. La normalización espacial no es algo trivial, de hecho para algunas escalas anatómicas no está claro siempre que la solución exista. La idea básica de la normalización espacial consiste en usar una máscara para definir el espacio estándar en el que diferentes sujetos se adaptan.

Mediante el uso de una plantilla que conforme el espacio de un sistema de coordenadas estándar, como el definido por Talairach y Tournoux [Talairach and Tournoux, 1988], es posible informar de posiciones anatómicas en términos de las coordenadas cartesianas, relativas a alguna referencia. Las transformaciones espaciales pueden dividirse en basadas en etiquetas y no basadas en las

mismas. Las primeras identifican características homólogas (etiquetas) en la imagen y plantilla y encuentran las transformaciones que mejor las superponen. Las aproximaciones en las que no se tiene en cuenta la etiqueta identifican una transformación espacial que minimiza algún índice de la diferencia entre un objeto y una plantilla, donde ambas se tratan como procesos continuos sin etiqueta [Toga, 1999].

Las deformaciones que se necesitan para transformar imágenes al mismo espacio no están definidas claramente, la tarea es menos acerca de la naturaleza de la transformación y más acerca de definir restricciones o elementos previos bajo los cuales una transformación es efectuada. La validez de la transformación se reduce generalmente a la validez de esta información previa, que se incorpora generalmente en forma de esquema bayesiano utilizando estimadores como máximo a posteriori (MAP, del inglés “maximum a posteriori”) o estimación de la varianza mínima (MVE, del inglés “minimum variance estimate”). Hay dos pasos involucrados para registrar imágenes conjuntamente. Está el registro en sí mismo, donde se determinan los parámetros que describen una transformación. Luego está la transformación, donde una de las imágenes se transforma de acuerdo a un conjunto de parámetros. El *algoritmo de optimización básico* tiene como objetivo determinar un conjunto de parámetros para los que alguna función de los parámetros se minimiza o maximiza. La optimización termina cuando se alcanza algún criterio de convergencia.

El proceso de la normalización espacial conlleva adaptar la imagen objeto a alguna forma de plantilla estandarizada. En la primera parte, se describen las transformaciones espaciales: en el primer paso de la normalización espacial, se determina un óptimo de 12 parámetros de transformación afín que registra las imágenes. Se continúa por un registro no lineal con el fin de corregir grandes diferencias que no fueron solventadas en el paso anterior de la normalización afín. En la siguiente sección se describe cómo las transformaciones de intensidad se pueden incorporar al registro.

Registro afín

Casi todos los corre registros entre sujetos o los métodos de normalización espacial para imágenes cerebrales comienzan por determinar los 9 o 12 parámetros óptimos de la transformación afín que registran las imágenes conjuntamente. Este paso generalmente se consigue minimizando o maximizando algunas funciones mutuas de las imágenes [Toga, 1999]. El objetivo del registro afín es adaptar la imagen \mathbf{f} a la plantilla \mathbf{g} usando una transformación afín de 12 parámetros (de p_1 a p_{12}). Las imágenes pueden ser escaladas de forma muy diferente, así que los parámetros de escala de la intensidad adicional (p_{13}) se in-

cluyen en el modelo. Hay muchas formas de parametrizar una transformación afín [Gispert et al., 2003]. Los parámetros más simples para optimizar son los elementos de la matriz de transformación. Un mapeado de transformación afín (via la matriz \mathbf{M} , donde los elementos de la matriz son los parámetros m_1 a m_{12} de la posición \mathbf{x} en una imagen a la posición \mathbf{y} en la otra se define por $\mathbf{y} = \mathbf{M} \cdot \mathbf{x}$ como a continuación:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} m_1 & m_4 & m_7 & m_{10} \\ m_2 & m_5 & m_8 & m_{11} \\ m_3 & m_6 & m_9 & m_{12} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix}$$

De acuerdo al algoritmo de optimización básico [Toga, 1999], los parámetros \mathbf{p} se pueden optimizar minimizando la suma de la diferencia de los cuadrados entre imágenes. La función minimizada es:

$$\sum_i f(M \cdot x_i) - p_{13} \cdot g(x_i)^2 \quad (3.10)$$

El registro afín puede hacer uso del algoritmo MAP, incorporando información previa en el procedimiento de optimización. La normalización espacial también puede ser no lineal, usando deformaciones que consisten en una combinación lineal de las funciones básicas. En la optimización, que consiste también en la minimización de la suma de las diferencias de los cuadrados entre la imagen objeto \mathbf{f} y la imagen plantilla \mathbf{g} . Las imágenes se pueden escalar de forma diferente, así que se necesita un parámetro adicional ω para acomodar estas diferencias:

$$\sum_i (f(y_i) - \omega \cdot g(x_i)^2) \quad (3.11)$$

Regularización

Sin regularización, es posible introducir deformaciones (en el registro afín) innecesarias que sólo reducen la suma residual de los cuadrados en una cantidad diminuta. Esto podría hacer el algoritmo muy inestable. La regularización se consigue minimizando la suma de los cuadrados diferencia entre la plantilla y la imagen deformada, mientras que simultáneamente se minimiza alguna función de las derivadas del campo de deformación [Gispert et al., 2003].

Los principios son bayesianos y hacen uso del esquema MAP incorporado a la optimización, que favorece una convergencia más rápida a una buena

solución y también aporta resultados más robustos y fiables para un número de datos limitado. El primer requerimiento para una aproximación MAP consiste en definir alguna forma de distribución a priori para los parámetros. Para una aproximación lineal simple, la distribución previa se puede basar en la media de los parámetros y también en la matriz de covarianza describiendo la distribución de los parámetros acerca de esta media, aunque hay muchas formas posibles para describirlos [Toga, 1999].

3.4.6 Normalización en intensidad

Las imágenes que nos ocupan en este trabajo son funcionales y por tanto no es posible llevar a cabo una comparación directa de intensidades de los vóxeles sin normalizar las intensidades, incluso para adquisiciones diferentes del mismo sujeto. La optimización al normalizar que se busca puede asumirse como la minimización de dos conjuntos de parámetros [Toga, 1999]: aquellos que describen las transformaciones espaciales (p_s), y los que describen las transformaciones de intensidad (p_t). Esto significa que la función diferencia puede expresarse de forma genérica como:

$$e_i(p) = f(s(x_i, p_s)) - t(x_i, p_t) \quad (3.12)$$

donde f es el objeto imagen, $s()$ es una función vector que describe las transformaciones espaciales basadas en los parámetros p_s y $t()$ es una función escalar que describe las transformaciones de intensidad basadas en los parámetros p_t . x_i representa las coordenadas del i -ésimo punto muestreado.

Los estudios de activación del cerebro implican identificar regiones de flujo cerebral incrementado mientras que los sujetos realizan diferentes tareas mentales en el escáner. En la mayoría de las modalidades de imagen funcional, los incrementos en el flujo sanguíneo conllevan incrementos en la intensidad de la imagen. Las variaciones globales pueden estar causadas por las diferencias en la dosis del trazador radiactivo, la sensibilidad del escáner, o la posición de la cabeza en el escáner. Las variaciones locales pueden ocurrir normalmente como diferencias entre pacientes, o con los mismos pacientes en momentos diferentes, y/o como resultados de la patología cerebral. Estas diferencias deben ser tenidas en cuenta para poder realizar comparaciones estadísticas, y se realiza escalando las imágenes entre sí o a algún estándar [Saxena et al., 1998].

Las intensidades en una imagen puede que no varíen linealmente con las intensidades en otra, así que puede ser más apropiado asociar una imagen a alguna función de otra imagen. Un ejemplo simple de esto podría ser adaptar

la imagen f a una versión escalada de la imagen g . Por ejemplo, podemos asumir un modelo polinomial para la transformación de la intensidad. Más variaciones complejas podrían ser incluidas modulando con otras funciones base, tales como la función básica DCT [Friston et al., 2007]. Por otro lado, [Collins et al., 1995] propuso que más que adaptar la imagen en sí misma a una plantilla, se podría utilizar transformaciones de magnitud de gradiente ya que son soluciones más robustas para solventar las no-homogeneidades en intensidad.

El análisis puede elevar falsamente la perfusión de áreas normales y disminuir la perfusión de áreas atróficas en el cerebro de los pacientes con AD cuando se usan métodos de escalado convencionales. SPM proporciona varios modelos para realizar el escalado, por ejemplo ANOVA asume una relación proporcional entre los cambios globales y regionales del rCBF, es decir, las activaciones son proporcionales al flujo global sanguíneo subyacente y la varianza se incrementa con el flujo sanguíneo global [Saxena et al., 1998]. El decremento en el flujo sanguíneo de la AD es suficientemente grande como para deprimir significativamente la media global de las intensidades de los vóxeles (rCBF equivalente) en el grupo de los controles, haciendo de ANOVA una selección pobre para el escalado, dado que asume que hay poca o ninguna varianza global. Esto provoca una pérdida de sensibilidad en las áreas más significativas para detectar la atrofia ocasionada en el AD. En [Saxena et al., 1998] se describe un proceso de normalización para solucionar esta pérdida de sensibilidad. En primer lugar, tras la normalización espacial anatómica estándar descrita por Talairach y Tournoux [Talairach and Tournoux, 1988], se aplicó un kernel de suavizado (FWHM $20 \times 20 \times 12$) a las imágenes normalizadas. Estas imágenes se escalaron por el procedimiento descrito a continuación:

1. El promedio de la intensidad de los vóxeles para cada imagen se calculó para los controles y AD. Se asume que los vóxeles con las intensidades mayores (que denotan los niveles más altos de perfusión) deberían estar en aquellas áreas que no estén afectadas por AD.
2. Cada vóxel en cada imagen se promedia por el 1% de la intensidad máxima de vóxel para esa imagen.
3. Para retener los dígitos más significativos, multiplicamos todos los vóxeles por 100.
4. Estas imágenes escaladas fueron proporcionadas por SPM [Stoeckel, 2003] para el análisis usando un diseño replicando condiciones de un sujeto. Las estadísticas utilizaron diferentes umbrales y un valor p de 0.05 que

se elige como estadísticamente significativo. Se experimentó con un 30% a un 0.25% de promedio. El promedio de 0.5% y 0.25% no mejoró significativamente los resultados comparados al 1% del promedio. Debido a esto y a que sospechamos que los outliers ejercen un efecto en los promedios tomados sobre un número pequeño de vóxeles, se tomó el 1% como estándar.

En el caso de esta tesis doctoral, y de una forma similar al método descrito en [Saxena et al., 1998], el nivel de intensidad de las imágenes se normaliza a la intensidad máxima que se calcula para cada imagen individualmente promediando sobre el 0.1% (3% para SPECT) de las intensidades de los vóxeles más altas. Este umbral se corresponde con el décimo bin en un histograma de intensidad de un total de 50 bins. Con este método evitamos tener en cuenta las regiones fuera del cerebro que suelen tener una baja intensidad previniendo una modificación del histograma de las imágenes y reduciendo así el efecto de posibles outliers de intensidad.

Capítulo 4

Diagnóstico asistido por computador

Este capítulo describe las herramientas CAD que son capaces de reproducir y complementar de forma automatizada el conocimiento de los médicos expertos en la evaluación de un paciente dado, por ejemplo para el diagnóstico precoz de la AD. El origen de su utilización radica en la necesidad de detectar objetivamente cambios mínimos que no se reconocen siempre mediante un diagnóstico meramente visual por parte de los médicos. A nivel de inferencia estadística, los sistemas CAD se clasifican en univariados y multivariados según si el análisis se realiza a nivel de vóxel o considerando grupos de vóxeles respectivamente. Dentro de los sistemas univariados, el paradigma SPM es capaz de medir las diferencias entre dos grupos realizando tests estadísticos, conocidos como mapas T o F resultado de comparar los valores de una imagen con los valores medios del grupo de controles. Por otro lado los sistemas multivariados definen los vectores de características que representan las diferentes imágenes funcionales para entrenar un clasificador automáticamente y distinguir entre controles y pacientes AD.

4.1 Introducción

Existen déficits mínimos de perfusión en las etapas tempranas de la AD, y cambios relacionados con la edad, que se ven frecuentemente en sujetos sanos y tienen que ser discriminados de los cambios específicos mínimos de la enfermedad [Benoit et al., 2002]. Estos cambios mínimos hacen del diagnóstico visual una tarea difícil que requiere médicos expertos. Con el fin de enfatizar la exactitud de predicción, especialmente en la etapa temprana de la AD, donde los pacientes se podrían beneficiar más de fármacos y tratamientos, son muy recomendables las herramientas CAD [Frackowiak et al., 2003]. El potencial de los sistemas CAD no se ha explotado totalmente en este área aunque se pueden encontrar algunas aproximaciones para la AD en la literatura [Fung and Stoeckel, 2007; Ishii et al., 2006; Minoshima et al., 1995].

El objetivo principal de los sistemas CAD es reproducir el conocimiento de los médicos expertos en la evaluación de un sujeto dado, por ejemplo para distinguir los pacientes AD de los controles. Por lo tanto, los errores en la evaluación de un único observador se pueden reducir usando un sistema de soporte de decisión (DSS, del inglés “decision support system”) [Sen and Biswas, 1985] para asistir a la identificación de los signos tempranos de la AD. En este sentido, se han propuesto algunas aproximaciones para analizar SPECT, PET y otras modalidades de imagen médica.

La estrategia común para enfrentarse al análisis de imagen médica se basa en la inferencia estadística, a nivel de vóxel (univariado) o considerando grupos de vóxeles o ROIs (multivariado) como explicamos en la sección 1.2.1 de esta tesis. La aproximación del análisis univariado más relevante a tener en cuenta es el ampliamente utilizado SPM y sus numerosas variantes [Friston et al., 2007].

4.2 Mapas estadísticos paramétricos (SPM)

SPM se utiliza generalmente para identificar respuestas cerebrales específicas y es la aproximación más prevalente para caracterizar la anatomía funcional y los cambios específicos de la AD [Ashburner, 2012], [Hu et al., 2005], [Ashburner, 2009]. Presentada en 1991 y desarrollada por el centro Trust Centre for Neuroimaging de Londres, se ha convertido en una referencia en el campo de la neuroimagen siendo ampliamente utilizada en departamentos como el de psiquiatría, medicina nuclear, radiología, farmacología, bioestadística, física médica etc. SPM permite el análisis de imágenes fMRI, PET, SPECT; así como EEG (Electroencefalografía, del inglés “Electroencephalography”) y MEG

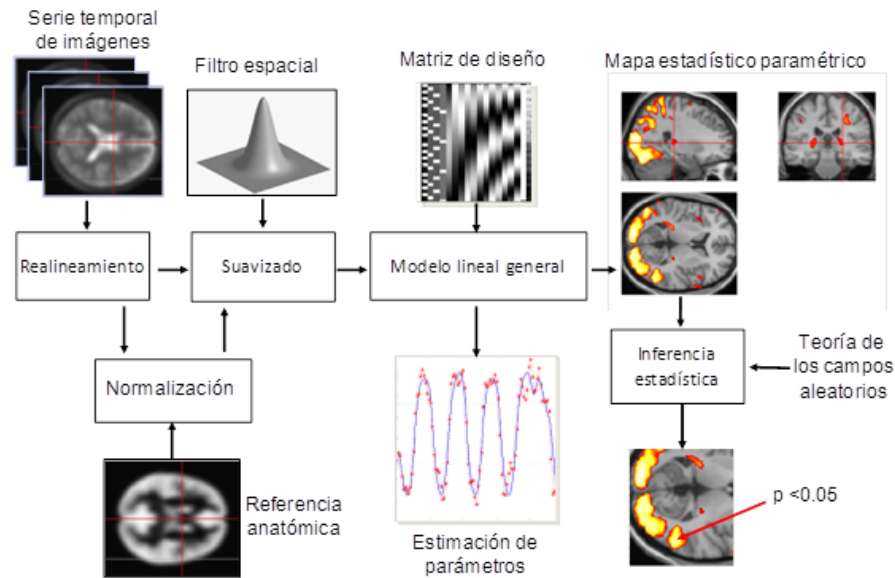


Figura 4.1: Etapas del análisis estadístico en SPM. Fuente [Friston et al., 2007]

(Magnetoencefalografía, del inglés “Magnetoencephalography”). Aunque es en la primera modalidad de imagen (fMRI) en la que SPM tiene mayor difusión por la facilidad que aporta para trabajar con secuencias de imágenes de un mismo paciente. El análisis de la base de datos de neuroimagen funcional involucra muchos pasos (como se representa en la figura 4.1) que se pueden dividir en [Friston et al., 2007]:

1. Preprocesamiento espacial
2. Estimación de los parámetros de un modelo estadístico
3. Realización de inferencias de la estimación de parámetros con las estadísticas apropiadas

La primera etapa de preprocesamiento comienza con una serie de transformaciones espaciales para reducir efectos no deseados de varianza entre vóxeles que se inducen por el movimiento del sujeto durante la adquisición o la forma cerebral que difiere entre una serie de escáneres. El primer paso consiste en realinear los datos para deshacer los efectos del movimiento del sujeto durante la sesión de escáner [Woods et al., 1992]. Después del realineamiento, los datos

se transforman usando deformaciones lineales o no-lineales en un espacio estadístico estándar. Finalmente, los datos se suavizan espacialmente antes de introducir la estadística apropiada.

4.2.1 Análisis estadísticos

En la segunda etapa, SPM conlleva la construcción de procesos estadísticos para testear las hipótesis de efectos específicos por regiones [Friston et al., 1991] que se engloban dentro del modelo lineal general (GLM, del inglés “General Linear Model”) [Chatfield and Collins, 1980]. La formulación del GLM se basa en dos conceptos: la matriz de diseño y los contrastes. Para el diseño del modelo, se pueden llevar a cabo dos clases de estudios: paramétricos o factoriales y categóricos o sustractivos. Los primeros estudian la relación entre las imágenes y un parámetro (como por ejemplo la edad o el resultado de un test cognitivo) y los segundos diferencian entre grupos mediante variables categóricas.

La matriz de diseño en los estudios paramétricos consta de una fila por cada imagen y de una columna por cada efecto, incluyendo los valores de los efectos correspondientes a cada sujeto. SPM es capaz de calcular la relación entre las imágenes y el parámetro linealmente. Las hipótesis de las respuestas no-lineales necesitan una transformación adecuada para linealizarlas correctamente. Por otro lado, en los estudios categóricos es necesaria una columna para determinar la pertenencia a cada uno de los grupos. Se pueden combinar ambos grupos de estudio (es decir, no es excluyente la pertenencia), también es posible modelar interacciones entre efectos, porque el efecto de un parámetro puede ser más intenso en un grupo que en otro.

Una vez establecido el modelo, SPM puede estimar la contribución de cada efecto de forma separada mediante los contrastes para diferenciar entre efectos de interés y efectos correctores [Friston et al., 2007]. Un contraste se define mediante un vector con longitud el número de efectos incluidos en la matriz de diseño. En el caso de los efectos categóricos, los contrastes deben cumplir una condición importante: la suma de todos los pesos en el contraste debe ser igual a cero. Si el efecto es corrector se pondera con un cero en el vector contraste. En caso de que el efecto sea paramétrico, el contraste determina si la corrección buscada es positiva, mediante un 1, o bien negativa mediante un -1, en la posición correspondiente a ese efecto en el vector contraste. En la figura 4.2, se presenta una matriz de diseño de un estudio categórico obtenida de la base de datos PET-Cartuja usada en este trabajo (véase la descripción de esta base de datos en el capítulo 8). Al contar con 2 efectos paramétricos, tenemos dos columnas que representan respectivamente la pertenencia al grupo

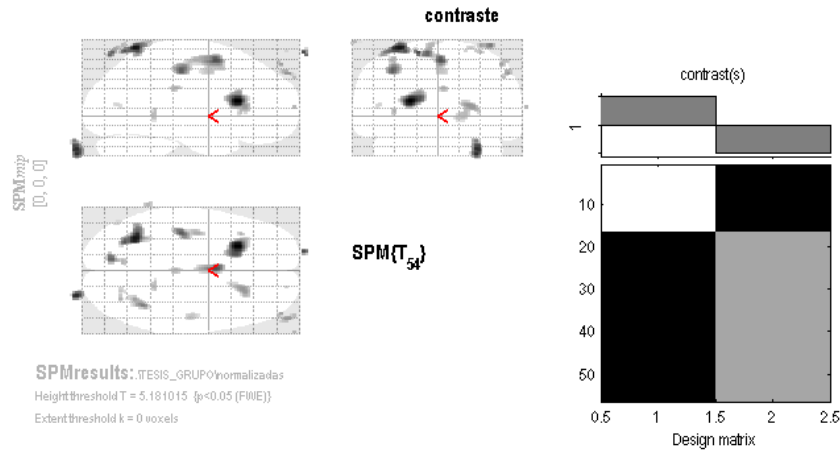


Figura 4.2: Vector de contrastes $[1 -1]$, matriz de diseño y resultados de SPM de un estudio categórico para la base de datos PET-Cartuja.

control y la pertenencia al grupo de AD. También se definen en esta figura los contrastes $[1 -1]$. La pertenencia al grupo de pacientes AD contribuye con un peso negativo (menor metabolismo que controles), mientras que el grupo de controles contribuye con un peso positivo (mayor metabolismo que pacientes). Si quisiéramos comprobar qué regiones presentan un metabolismo mayor en AD, el contraste sería $[-1 1]$.

4.2.2 Obtención de los SPMs

En la tercera etapa, se obtienen los mapas estadísticos paramétricos (o SPMs), que son procesos de imagen con valores de vóxel (llamados p) que se distribuyen bajo la hipótesis nula de acuerdo a una probabilidad de función de densidad conocida, generalmente las distribuciones T o F de Student, conocidas coloquialmente como mapas T o F.

Los valores de p se interpretan a priori con un umbral [Chatfield and Collins, 1980] que suele ser $p < 0.05$ cuyo significado es el de admitir un falso positivo cada 20 observaciones, aunque cada vez es más frecuente realizar tests de permutación con remuestreo para establecer y validar los umbrales reales de significación. Para interpretar los valores p se distingue entre niveles de evidencia ordenados de mayor a menor fiabilidad: los niveles de evidencia 1 y 2 corresponderían a valores de p corregidos menores que $p = 0.05$, mientras que en casos de p sin corregir inferiores a $p = 0.001$ (nivel de evidencia 3), se

necesita un soporte bibliográfico o hipótesis biológica que los avale, ya que es un umbral mucho menos estricto que $p = 0.05$. En la figura 4.2 se observa también el resultado de SPM de comparar pacientes con AD y sujetos sanos, en el que se reflejan las zonas de disminución de rCBF de pacientes con AD [Matsuda et al., 2007]. Consideramos el umbral $p = 0.05$ corregido para la obtención de la figura.

SPM es capaz de medir las diferencias entre dos grupos haciendo este test estadístico a nivel de vóxel, por ejemplo en un t-test de dos muestras, comparando los valores de una imagen de un paciente (normal o AD) con los valores medios del grupo de controles. A continuación, los vóxeles significativos son inferidos usando la *teoría de campos aleatorios* [Adler, 1981]. Se destaca también que los SPMs se interpretan como procesos estadísticos extendidos espacialmente mediante la referencia al comportamiento probabilístico de los campos gaussianos para tratar adecuadamente el problema de la no-independencia de vóxeles, sobre todo los cercanos.

4.3 Aproximaciones multivariadas

Las aproximaciones multivariadas [Meda et al., 2012], [Modinos et al., 2012], [Hansen, 2007] consideran los vóxeles como niveles diferentes de un experimento o factor de tratamiento y las secuencias de datos de todos los vóxeles juntos, frente a SPM que lo hace vóxel a vóxel. Asimismo las aproximaciones multivariadas se dividen en aquellas de conectividad efectiva (que inferen por naturaleza) y en aquellas de conectividad funcional o patrones de covarianza (que son dirigidas o exploratorias) [Friston et al., 2007]. En lo concerniente a las aproximaciones multivariadas de conectividad efectiva, los modelos lineales asumen que las entradas múltiples a una región cerebral son linealmente separables. Esta suposición impide las conexiones dependientes de actividad que se expresan en un contexto. El problema se resuelve adoptando modelos que incluyen interacciones entre entradas. Estas interacciones o efectos bilineales pueden construirse como un contexto o actividad de modulación dependiente de la influencia que una región ejerce sobre otra, donde ese contexto está ejemplificado por la actividad en regiones cerebrales más lejanas ejerciendo efectos moduladores. Estas no linealidades pueden introducirse al modular usando las llamadas variables moderadoras que representan la interacción entre dos regiones causando actividad en una tercera.

Si consideramos las aproximaciones multivariadas que se basan en la conectividad funcional o patrones de covarianza, la mayoría de los análisis de covarianza entre regiones cerebrales se basan en la decomposición del valor singular

[Friston et al., 2007] (SVD, del inglés “Singular Value Decomposition”) o en el análisis de componentes principales (PCA) que representan con eigenimágenes modos espaciales de actividad cerebral coherente. El análisis de eigenimágenes se ha elaborado de numerosas formas. Se destaca el análisis variado canónico (CVA, del inglés “canonical variate analysis”) y el escalado multidimensional. CVA fue introducido en el contexto del análisis múltiple de covarianza (MANCOVA, del inglés “Multiple Analysis of Covariance”) y que hace uso de la solución general del eigenvector para maximizar la varianza que puede ser explicada mediante variables explicativas relacionadas con el error. CVA se puede considerar como una extensión del análisis de eigenimagen que se refiere explícitamente a algunas variables explicativas y permite la inferencia estadística. Las aproximaciones multivariadas como MANCOVA consideran todos los vóxeles en un único escáner como una observación para realizar inferencias acerca de los efectos de activación distribuidos. Su importancia se basa en el hecho de que los efectos debidos a activaciones son evaluados estadísticamente en términos de los efectos en cada vóxel y también interacciones entre vóxeles, pero requiriendo un número de observaciones (escáneres) mayor que el número de componentes de la observación multivariada (es decir, vóxeles) [Frackowiak et al., 2003].

Hay otras aproximaciones importantes como el análisis de componentes independientes (ICA) [McKeown et al., 1998] que utiliza la maximización de la entropía para conseguir (usando esquemas iterativos) modos espaciales o sus dinámicas que sean aproximadamente independientes.

Finalmente ha habido un interés importante en el análisis de agrupaciones (del inglés “clustering”) [Baumgartner et al., 1997]. Conceptualmente, esto se puede relacionar con el análisis de eigenimagen a través de escalado multidimensional y análisis de coordenadas principales. En el análisis de agrupaciones se caracterizan los centroides de la agrupación y los modos espaciales definidos por la probabilidad de pertenecer a cada cluster. La figura 4.3 muestra un ejemplo de “clustering” basado en mezcla de gaussianas [Segovia et al., 2010]. Otra aproximación prolífica al problema de análisis de imagen médica se basa en la teoría de aprendizaje estadístico [Toivonen, 1996; Vapnik, 1998]. Un ejemplo es la propuesta simple vóxeles como características (VAF, del inglés “Voxels as Features”) que se representa en la figura 4.4, en la que las intensidades de los vóxeles I_n de la imagen se utilizan directamente para construir los vectores de características $v = (I_1, \dots, I_N)$ [Stoeckel et al., 2001], [Stoeckel et al., 2004]. Este esquema describe el conocido problema de la *maldición de las dimensiones* [Duin, 2000] que es muy común en los estudios de imagen funcional dado que el número de imágenes es limitado en comparación con el ele-

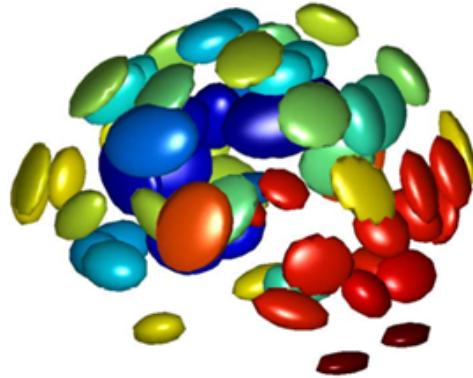


Figura 4.3: Método de “clustering” basado en mezcla de gaussianas. Fuente: [Segovia et al., 2010]

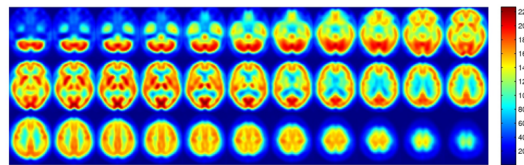


Figura 4.4: VAF para base de datos SPECT: Los valores de todos los voxels son utilizados para construir el vector de características

vado número de características (500000 vóxeles para la base de datos SPECT de la figura 4.4). Se han propuesto algunas soluciones para solventar este problema dentro del contexto de aproximaciones multivariadas *supervisadas* (se tiene en cuenta la etiqueta de cada paciente frente a las no-supervisadas que no lo hacen), como en [Ramírez et al., 2009a], [Álvarez et al., 2009], [López et al., 2009b], [Salas-Gonzalez et al., 2010b], [Illán et al., 2011b]. En primer lugar, se definen los vectores de características que representan las diferentes imágenes funcionales. En segundo lugar, un clasificador se entrena usando un conjunto dado de muestras conocidas [Stoeckel et al., 2001], [Stoeckel et al., 2004], [Fung and Stoeckel, 2007] para distinguir automáticamente entre controles y pacientes AD.

Una aproximación de aprendizaje estadístico conforma una herramienta potente para el análisis de imagen cerebral y de esta forma ayudar en la toma de decisiones, con la ventaja adicional de que no es necesario un conocimiento específico de la enfermedad. En este tipo de aproximaciones aunque las

imágenes son etiquetadas por médicos expertos (teniendo en cuenta también los test clínicos), evitamos la evaluación convencional que a menudo confía en una reorientación manual, una lectura visual de los cortes tomográficos y un análisis semicuantitativo de ciertas regiones de interés. Estos pasos consumen tiempo, son subjetivos y propensos a error [Silverman et al., 2001].

Además, la memoria de largo plazo se ubica en las cortezas sensoriales de alto orden o secundarias y de asociación [Cardinali, 2007]. Las cortezas sensoriales de primer orden (también conocidas como “primarias”) distribuyen información a las mencionadas cortezas de asociación o de alto orden [Contreras et al., 2008], las cuales permiten el reconocimiento de parte del mundo sensorial, como por ejemplo las dos áreas corticales que participan en el reconocimiento visual de caras. Estas cortezas de asociación son las que envían información a las cortezas prefrontales ejecutivas, que incluyen la corteza prefrontal medial, la orbitofrontal y la cingulada anterior. En este sentido, esta tesis propone un sistema CAD multivariado para mejorar la detección precoz de la AD que se centra en el análisis de la imagen, señalando la importancia de las relaciones de alto orden entre áreas cerebrales mediante ARs (que no se pueden encontrar directamente a nivel de vóxel o en el análisis de regiones de interés).

Parte III

Fundamento teórico

Capítulo 5

Fundamentos teóricos

En este capítulo introducimos los fundamentos teóricos relativos a los métodos de selección, extracción de características, clasificación y evaluación de un sistema CAD. Respecto a los métodos de selección de características, se propone el criterio FDR para la obtención de vóxeles discriminantes. Como eje principal de este trabajo, se describe formalmente el problema relacionado con las reglas de asociación y los algoritmos para su extracción, dado que pueden utilizarse como método de selección de características. También se introducen los métodos de discretización puesto que las ARs pueden establecerse no sólo entre atributos continuos o discriminantes, sino también entre atributos discretos. En lo que a los métodos de extracción de características se refiere, asimismo se describen los métodos clásicos de extracción basados en PCA y PLS que reducen el espacio de entrada en base a la varianza de los datos. A continuación se expone el algoritmo LMNN como novedoso extractor y transformador del espacio de características seleccionadas, aplicado a imagen funcional en esta Tesis. Por último se explica la finalidad y fundamento de los clasificadores, incidiendo en los que aporta este trabajo: el novedoso modelo de clasificación basado en ARs y el clasificador LMNN (el cual se compara con SVM dadas sus similitudes). Finalmente se presentan los métodos de evaluación del rendimiento del sistema CAD, entre ellos el análisis ROC y AUC.

5.1 Métodos de selección de características

La alta dimensionalidad del espacio de entrada de características en comparación con el reducido número de sujetos [Costafreda et al., 2009] es un problema ampliamente conocido como *Maldición de la dimensión*, que se resuelve a menudo con técnicas de selección de características (FS, del inglés “Feature Selection”). La FS tiene beneficios porque acelera el proceso de test e incrementa la exactitud de clasificación [Duchesnay et al., 2011], [Polat and Güneş, 2009]. El desarrollo de la selección de características se orienta en dos direcciones fundamentalmente: la selección de subconjuntos y el rango de las características. La principal dificultad para selección de conjuntos de características, radica en que el aprendizaje estadístico se enfrenta con el problema de seleccionar un subconjunto relevante de características al cual hay que prestar atención e ignorar el resto [Kohavi and John, 1997].

La selección de subconjuntos se clasifica en tres categorías: filtros, envoltorios y embebidos [Maldonado et al., 2011]. El método de filtrado usa las propiedades estadísticas de las características para *filtrar* las que aportan más información [Liu et al., 1998a]. Las aproximaciones de filtro puntúan y categorizan las características basándose en ciertos criterios estadísticos seleccionando las características con los valores de puntuación más altos. Los métodos de filtro más frecuentemente utilizados son: t-test, test chi-cuadrado, test de Mann-Wilcoxon Whitney, información mutua, coeficientes de correlación de Pearson o PCA [Gheyas and Smith, 2010]. Los métodos de filtrado son rápidos pero con falta de robustez frente a las interacciones entre características y redundancia de las mismas. Además no está claro cómo determinar los puntos de corte para establecer las puntuaciones de las características realmente importantes y excluir las superfluas.

El método de envoltorio explora todo el conjunto de características para puntuar los subconjuntos de características de acuerdo a su potencialidad predictiva, optimizando un criterio de rendimiento del consiguiente algoritmo para la clasificación [Kohavi and John, 1997]. Es decir, la selección de características *se envuelve* en un algoritmo de aprendizaje que se aplica a subconjuntos de características y se prueba con validación verdadera (un conjunto de observaciones se elige aleatoriamente como test y el resto de las observaciones se utilizan en el entrenamiento) obteniendo una exactitud para determinar la calidad del conjunto de características. Dado que la búsqueda exhaustiva no es computacionalmente fiable, los métodos de envoltorio deben emplear un algoritmo de búsqueda para encontrar un conjunto óptimo de características.

Los métodos de envoltorio se pueden clasificar en dos categorías basadas en estrategias de búsqueda: voraz y aleatorizado/estocástico. Los métodos

de envoltorio voraz utilizan menos tiempo computacional que otras aproximaciones de envoltorio y se clasifican en [Kohavi and John, 1997] selección secuencial hacia atrás (SBS, del inglés “Sequential Backward Selection”) y selección secuencial hacia delante (SFS, del inglés “Sequential Forward Selection”) que son los dos métodos más comúnmente utilizados de envoltorio que utilizan una estrategia de búsqueda voraz denominada *escalada de montaña*. SBS comienza con el conjunto de todas las características y progresivamente elimina las menos prometedoras. SBS para sí el rendimiento de los algoritmos de aprendizaje cae por debajo de un umbral dado debido a la eliminación de las características restantes. SBS asume y confía en el comportamiento monótono ya que conforme el número de características aumenta, la exactitud de la predicción no decrece. Esta suposición es discutible debido a las dificultades asociadas con la dimensionalidad del espacio de búsqueda y el sobreentrenamiento. En realidad, la habilidad predictiva de un algoritmo de aprendizaje puede decrementarse conforme la dimensión del subespacio de características aumenta después de alcanzar un punto máximo. Cuando nos encontramos con datos de gran dimensionalidad, SBS a menudo encuentra dificultades para identificar el efecto separado de cada variable explicativa en la variable objetivo. Debido a esto, los buenos predictores pueden eliminarse temprano en el algoritmo (en SBS, una vez que una característica es eliminada, es eliminada de forma permanente).

Como contraste, SFS comienza con un conjunto vacío de características y selecciona una característica en cada iteración, comenzando con la característica más prometedora hasta que no haya mejora de exactitud en la clasificación. En SFS, una vez que se añade la característica nunca se elimina. SBS y SFS pueden verse atrapados en un mínimo local debido a la búsqueda de seguimiento único que realizan. Aunque se han sugerido métodos de búsqueda flotante (SFFS, SFBS) que secuencialmente añaden o quitan características, pero que tienen como principal problema que la utilidad de la característica individual no es a menudo suficiente por sí misma siendo necesario incluir las combinaciones de otras características. Actualmente, el algoritmo de aprendizaje más popular usado en los esquemas de envoltorio es el de SVM [Bishop, 2006]. Sin embargo, la exactitud de un SVM depende de la elección del kernel y de los parámetros (por ejemplo, parámetros de coste, variables de pérdidas, margen del hiperplano etc). El error al encontrar los parámetros SVM puede afectar a la exactitud de predicción. Otro punto flaco es que computacionalmente demandan más que los métodos de filtro porque evalúan los conjuntos de características usando un algoritmo de aprendizaje, y estos suelen ser usualmente iterativos. Para acelerar la aproximación de envoltorio en la búsqueda

es necesario utilizar un algoritmo de aprendizaje rápido. Además hay una evidencia empírica que sugiere que los SVMs son muy sensibles a los datos de entrenamiento ruidosos, que pueden degradar su rendimiento. El desarrollo de algoritmos de búsqueda rápida y con exactitud para la selección de subconjuntos óptimos es un tema abierto a mejoras.

Por otro lado, recientemente algunos autores han propuesto soluciones híbridas (entre filtros y envoltorios). Esta aproximación (métodos embebidos) lleva a cabo la selección de características en el proceso de construcción del modelo. La idea que subyace es que primero se apliquen los métodos de filtrado para seleccionar un conjunto de características y a continuación se utilice el método de envoltorio para seleccionar las características óptimas a partir de las anteriores. Los métodos híbridos argumentan que el riesgo de eliminar buenos predictores mediante filtros desaparece si su punto de corte para un listado ordenado de características es bajo. Sin embargo, los métodos híbridos pueden sufrir pérdidas de exactitud dado que un conjunto relevante de características aislado puede aparecer no más discriminante que uno irrelevante en presencia de interacciones de características. Algunos ejemplos de métodos híbridos se encuentran en [Gheyas and Smith, 2010]. Finalmente, en lo que a metodología de rango de características se refiere, el método hace uso de información de rango en lugar de verlo simplemente como una categoría plana. El objetivo final es encontrar un conjunto de características con la mayor importancia y la menor similitud, es decir, intenta evitar seleccionar características redundantes dando lugar a un algoritmo de búsqueda ambicioso para resolver problemas de optimización [Geng et al., 2007]. En este trabajo se proponen las ARs como método de selección de características en imágenes tomográficas funcionales en el capítulo 6. También se propone el criterio FDR, para la selección de las coordenadas que sitúan los vóxeles más relevantes como se explica a continuación.

5.1.1 Criterio FDR

El análisis discriminante motiva la siguiente pregunta: dado un conjunto con dos clases, decir cuál es la mejor característica o conjunto de características (tanto lineales como no lineales) para discriminar entre las dos [Mika et al., 1999]. Las aproximaciones clásicas se enfrentan a esta pregunta comenzando con el (teóricamente óptimo) clasificador de Bayes, asumiendo distribuciones normales para las clases, así como algoritmos estándares entre los que se destaca el análisis discriminante cuadrático o lineal, de los que puede ser derivado el criterio discriminante lineal de Fisher [Fisher, 1936].

Criterio discriminante lineal de Fisher

Definimos $X_1 = \{x_1^1, \dots, x_{l_1}^1\}$ y $X_2 = \{x_1^2, \dots, x_{l_2}^2\}$ como las muestras de dos clases diferentes y $X = X_1 \cup X_2 = \{x_1, \dots, x_l\}$ [Mika et al., 1999]. El criterio discriminante lineal de Fisher está dado por el vector w que maximiza:

$$J(w) = \frac{w^T \cdot S_B \cdot w}{w^T \cdot S_w \cdot w} \quad (5.1)$$

donde:

$$S_B = (m_1 - m_2) \cdot (m_1 - m_2)^T \quad (5.2)$$

$$S_w = \sum_{i=1,2} \sum_{x \in X_i} (x - m_i) \cdot (x - m_i)^T \quad (5.3)$$

que son respectivamente las matrices de dispersión entre clase (S_B) y dentro de clase (S_w) y m_i es la media. Intuitivamente maximizar $J(w)$ consiste en encontrar una dirección que maximice las medias de la clase proyectada (en el numerador) mientras que se minimice la varianza de clases en esta dirección (en el denominador).

Umbral FDR para la selección de vóxeles

El criterio FDR es una de las técnicas lineales estándares en el análisis estadístico de datos y se caracteriza por su habilidad de separación entre clases [Webb, 2002]. Para el caso de dos clases, se puede definir como:

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (5.4)$$

donde μ_i y σ_i^2 denotan la i -th media y varianza de clase para cada variable de entrada (vóxel), respectivamente. Para un vóxel dado, el valor de ratio crece conforme la diferencia de los valores medios de las clases aumenta o la dispersión acumulada en cada clase decrece, siendo por lo tanto útil para descubrir los vóxeles discriminativos. Esta aplicación se ha utilizado en el caso de las imágenes funcionales [Lukic et al., 2002], para la selección de los vóxeles que satisfacen un nivel de umbral particular de FDR, identificándolos como las variables más discriminativas [Ramírez et al., 2009a], [Illán et al., 2011b]. Así en este trabajo el FDR se utilizó como método para la obtención de máscaras que contienen los vóxeles más discriminativos para su posterior clasificación (como se explica en el capítulo 6) [Chaves et al., 2012c].

TID	Items
1	{ <i>pan, leche</i> }
2	{ <i>pan, leche, cerveza, huevos</i> }
3	{ <i>leche, pañales, cerveza, cola</i> }
4	{ <i>pan, leche, pañales, cerveza</i> }
5	{ <i>pan, leche, pañales, cola</i> }

Tabla 5.1: Ejemplo de las transacciones para el estudio de un problema de cesta de la compra. Fuente: [Tan et al., 2005]

5.2 Reglas de asociación

En el área de minería de datos, existe un interés creciente en la técnica de las reglas de asociación para identificar tendencias y asociaciones de los datos que van más allá de un análisis simple [Menzies and Hu, 2003]. Este problema fue formulado inicialmente por Agrawal ([Agrawal et al., 1993], [Srikant and Agrawal, 1995]) y se conoce también como *el problema de la cesta de la compra*, dado que se originó del estudio de la relación en el consumo de productos en un mercado. En este problema, existe un conjunto de items y una gran colección de transacciones que son los subconjuntos (o cestas) de estos items. El análisis de datos mediante transacciones se usa comúnmente para mejorar la calidad de las decisiones. La tarea consiste en encontrar relaciones entre varios items dentro de los subconjuntos [Brin et al., 1997]. La siguiente tabla 5.1 ilustra un ejemplo de los datos conocidos como transacciones (o compras), que contienen un único identificador por transacción (TID, del inglés “Transaction Identifier”). Cada fila en esta tabla se corresponde con una transacción, que contiene un TID y un conjunto de items comprados por un cliente dado.

En esta sección se presenta una metodología conocida como análisis de asociación, que es útil para descubrir relaciones interesantes, y que están ocultas en grandes conjuntos de datos. Las relaciones descubiertas se pueden representar en la forma de ARs o conjuntos de items frecuentes. Por ejemplo, la siguiente regla se puede extraer del conjunto de datos mostrados en la tabla 5.1:

$$\{\textit{pañal}\} \rightarrow \{\textit{cerveza}\}$$

Esta regla sugiere que existe una fuerte relación entre la venta de pañales y cerveza porque muchos clientes que compraron pañales también compraron cerveza. Los analistas de mercados pueden utilizar estos datos para promocionar el mercado por ejemplo con el simple hecho de la colocación a partir del conocimiento de los items que son vendidos conjuntamente. Aparte del análisis

de mercados para estudiar la venta de productos, las ARs se pueden aplicar en otros dominios como, la bioinformática, el diagnóstico médico, el análisis de datos científicos, test de fraudes estudiados por compañías de seguros, predecir los resultados de un test por los resultados dados por otro test en investigación médica para evitar pruebas redundantes y las taxonomías para evitar reglas redundantes, estudios entre interacciones de diferentes elementos en el planeta Tierra etc. Al mismo tiempo cuando se extraen ARs hay principalmente dos problemas a los que enfrentarse. En primer lugar se encuentra la complejidad algorítmica ya que el número de reglas crece exponencialmente con el número de items. En segundo lugar, el hecho de que las reglas interesantes deben filtrarse del conjunto total de reglas generadas dado que existen reglas espúreas que pueden ocurrir simplemente por casualidad [Hipp et al., 2000].

5.2.1 Descripción formal del problema

Sea $I = \{x_1, \dots, x_n\}$ un conjunto de distintos literales, llamados items. Un conjunto $X \subseteq I$ se llama conjunto de items (en inglés “itemset”). Sea D una base de datos de transacciones de los items pertenecientes a I . Cada $T \in D$ se denomina transacción. Decimos que una transacción $T \in D$ soporta un conjunto de items $X \subseteq I$ si $X \subseteq T$ [Hipp et al., 2000]. Asociado con una transacción hay un único identificador, denominado su TID. Cada itemset posee un determinado número de items entre los que establecer ARs, por ejemplo un itemset de 3 items se denomina 3-itemset y en general si posee k items, se denomina “ k -itemset”. El conjunto vacío es el que no contiene ningún item.

Definición de AR

Una AR es una expresión que relaciona unidireccionalmente los itemsets X e Y que poseen intersección nula ($X \cap Y = \emptyset$), de la forma $X \implies Y$. Al mismo tiempo el significado intuitivo de una regla es que las transacciones en la base de datos que contienen los items en X tienden también a contener los items en Y . La fuerza de una AR se puede medir en términos de su soporte y confianza. El soporte determina con qué frecuencia se aplica una regla a un conjunto de datos dados, mientras que la confianza determina con qué frecuencia los items que aparecen en Y aparecen en las transacciones que contienen X [Agrawal et al., 1993].

Soporte y confianza de una AR

Una propiedad importante de un itemset es su soporte, que se refiere al número de transacciones que contienen un conjunto particular de items.

Matemáticamente, el soporte $\sigma(X)$, para un conjunto de items X puede establecerse como [Tan et al., 2005]:

$$\sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}| \quad (5.5)$$

donde el símbolo $|\cdot|$ simboliza el número de elementos en un conjunto. Por ejemplo en la tabla 5.1, la cuenta de soporte para $\{cerveza, pañales, leche\}$ es igual a dos porque sólo hay dos transacciones que contienen los tres items. De ahí que la regla $X \Rightarrow Y$ tiene un soporte s si $s\%$ de las transacciones en T (de un total de N transacciones) contienen $X \cup Y$ y una confianza c , si el $c\%$ de las transacciones en T que contienen X también contienen Y , [Wu et al., 2004] es decir:

$$soporte(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (5.6)$$

$$confianza(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (5.7)$$

El soporte es una medida importante dado que una AR que tiene bajo soporte puede ocurrir simplemente por casualidad y no ser interesante. La confianza, por otro lado, mide la fiabilidad de la inferencia establecida por una regla. Para una regla dada $X \Rightarrow Y$, cuanto mayor es la confianza, más probable es para Y estar presente en las transacciones que contienen X . La confianza también proporciona una estimación de la probabilidad condicional de Y dada por X .

El análisis asociativo debe interpretarse con cuidado, dado que la inferencia establecida mediante una regla de asociación, no implica necesariamente causalidad. En su lugar, sugiere una fuerte relación de co-ocurrencia entre items en el antecedente y consecuente de la regla. Sin embargo, la causalidad requiere conocer los atributos de causa y efecto en los datos y típicamente requieren relaciones en el tiempo (por ejemplo, la reducción del ozono implica el calentamiento global) [Tan et al., 2005].

Representación binaria

Los datos de mercado se pueden representar en un formato binario como se muestra en la tabla 5.2, donde cada fila se corresponde con una transacción y cada columna se corresponde con un item. Un item puede ser tratado como una variable binaria cuyo valor es “1” si el item está presente en una transacción y “0” en caso contrario. Debido a que la presencia de un item en una transacción se considera a menudo como más importante que su ausencia, un item es una variable binaria asimétrica. En esta tesis hemos adaptado esta representación

TID	Pan	Leche	Pañal	Cerveza	Huevos	Cola
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	0	1

Tabla 5.2: Representación binaria de imágenes funcionales. Fuente [Tan et al., 2005]

binaria al análisis de imágenes tomográficas funcionales mediante la elección de regiones activadas (“1”) o no activadas (“0”), tal y como se explica en el capítulo 6.

5.2.2 Algoritmos para la extracción de ARs

Para resolver el problema de la extracción de ARs, se dispone de un amplio número de aproximaciones descritas en numerosas referencias bibliográficas como [Hipp et al., 2000], [Agrawal et al., 1993], [Houtsma, 1995], [Borgelt, 2005], [Wang et al., 2005], [Kuo et al., 2011] que admiten la clasificación, dependiendo de un criterio basado en la estrategia para atravesar el espacio de búsqueda (como en los algoritmos de búsqueda primero en amplitud o BFS, del inglés “Broad First Search”) o para determinar los valores de soporte de los itemsets (como en los algoritmos de búsqueda primero en profundidad o DFS, del inglés “Depth First Search”). La figura 5.1 muestra las aproximaciones más comunes para los algoritmos de extracción de reglas, que se clasifican en en BFS y DFS. Con BFS, los valores de soporte de todos los (k-1)-itemsets se determinan antes de contar los valores de soporte de los k-itemsets. Sin embargo, DFS desciende recursivamente siguiendo una estructura en árbol Hipp et al. [2000]. Tanto BFS como DFS se subdividen a la vez en algoritmos de recuento e intersección según la forma de determinar el soporte de un itemset. Mientras en los primeros se cuentan las ocurrencias en la base de datos, en los segundos se determinan los valores de soporte de los candidatos mediante intersecciones de conjuntos.

Dentro de la clasificación BFS y los algoritmos de recuento, el algoritmo más popular de este tipo es *Apriori* [Agrawal et al., 1993] que descarta aquellos conjuntos candidatos que tienen un subconjunto no frecuente antes de contar sus soportes. Esta optimización es posible porque BFS asegura que los valores de soporte de todos los subconjuntos de un candidato se conocen

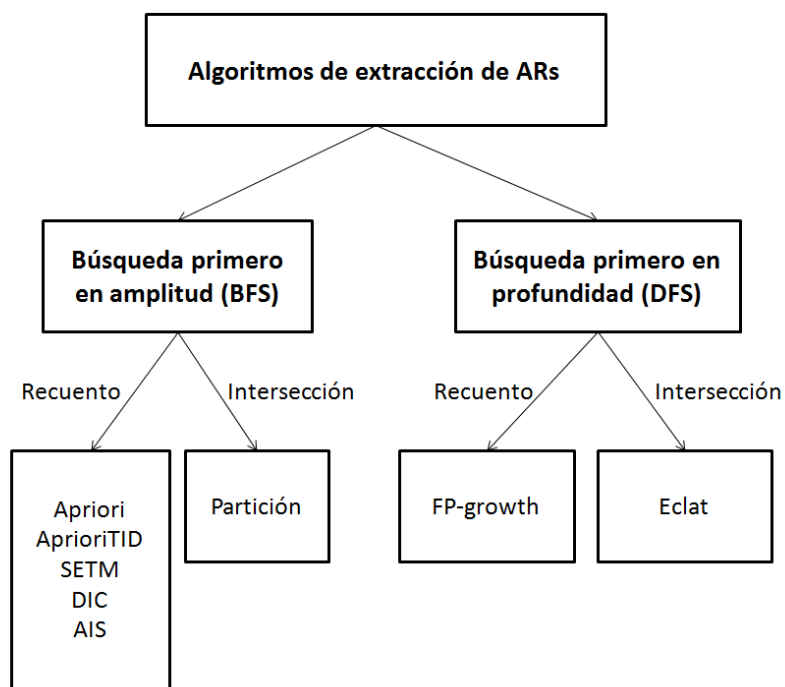


Figura 5.1: Algoritmos más utilizados para la extracción de ARs. Obtenido de [Hipp et al., 2000].

previamente. Por otro lado, *AprioriTID* [Srikant and Agrawal, 1995] es una extensión de la aproximación básica *Apriori* que en lugar de actuar sobre la base de datos sin procesar, representa internamente cada transacción mediante los itemsets candidatos (seleccionados previamente porque superan un soporte establecido) actuales que contiene. Con *AprioriHybrid* [Srikant and Agrawal, 1995] se combinan ambas aproximaciones (*Apriori* y *AprioriTID*). Por otro lado, *SETM* [Houtsma, 1995] surgió por la motivación de SQL para tratar computacionalmente conjuntos de items frecuentes. *SETM* consiste en un sólo bucle en el que se llevan a cabo dos operaciones de ordenación y una de unión de escáneres. La primera se necesita para implementar la unión de escáneres que la sigue. La segunda se usa para estimar el soporte eficientemente.

Las diferencias más importantes de *Apriori* con *AIS* y *SETM* se encuentran en la forma de producir y seleccionar conjuntos de items candidatos. En *AIS* y *SETM*, se generan candidatos sobre la marcha basándose en las transacciones leídas de la base de datos [Agrawal et al., 1993]. En *AIS*, los conjuntos de items numerosos se amplían con otro ítem individual en la transacción para generar conjuntos de items candidatos. Sin embargo, estos conjuntos de items pueden no ser numerosos, produciendo muchos conjuntos de items candidatos. El algoritmo *Apriori* resuelve este problema ya que une los conjuntos de items frecuentes y elimina los subconjuntos pequeños en la pasada anterior sin considerar la transacciones en la base de datos, reduciendo de este modo el conjunto de items candidatos frecuentes [Zubi and Saad, 2011] y acelerando el proceso.

El algoritmo *DIC* (en inglés, “Dynamic itemset counting”) [Kuo et al., 2011] es una variación mayor del algoritmo *Apriori*. *DIC* suaviza la separación estricta entre el cálculo y la generación de candidatos. Siempre que un candidato alcance un soporte mínimo, es decir incluso cuando este candidato no ha sido comprobado todavía en todas las transacciones, *DIC* comienza a generar candidatos adicionales basándose en ello. Para ese propósito se emplea un árbol de prefijos [Hipp et al., 2000].

Dentro de BFS los algoritmos de intersección (que consisten en el uso de conjuntos de intersecciones para determinar valores de soporte al igual que *Apriori*) destacamos el algoritmo de *partición* [Wang et al., 2005]. El algoritmo de *partición* pretende usar las listas de identificadores de transacción (“tidlists”) de los k-candidatos. Obviamente el tamaño de los resultados intermedios crece fácilmente más allá de las limitaciones de la memoria física de los ordenadores convencionales. Para resolver este inconveniente, el algoritmo de *partición* separa la base de datos en trozos que son tratados de forma

independiente, siendo ésta la principal diferencia con respecto a *Apriori*.

Los algoritmos de búsqueda de profundidad (DFS) definen sus conjuntos de candidatos como grupos de items de uno de los nodos del árbol (Un árbol es una estructura no lineal en la que cada nodo puede apuntar a uno/varios nodos. Los nodos son registros formados por un campo de datos y un número de apuntadores o enlaces [Summa et al., 2011]). La dificultad reside en el coste computacional que supone escanear la base de datos para cada nodo. Recientemente se ha introducido el algoritmo *FP-growth* [Borgelt, 2005]. La idea básica de *FP-growth* puede describirse como un esquema de eliminación recursiva: en el paso de preprocesado se borran todos los items de las transacciones que no son frecuentes individualmente, es decir que no aparecen en un número mínimo de transacciones especificadas por el usuario. Luego se seleccionan todas las transacciones que contienen los items menos frecuentes entre aquellos que son frecuentes y se borran. En contraste con las aproximaciones DFS básicas, *FP-growth* no sigue los nodos del árbol sino que directamente desciende a alguna parte del conjunto de items en el espacio de búsqueda. En un segundo paso, *FP-growth* utiliza un árbol FP para derivar los valores de soporte de todos los conjuntos de items frecuentes.

Dentro de DFS cuando se consideran intersecciones de las “tidlists”, se introduce el algoritmo *Eclat* que emplea una optimización denominada intersecciones rápidas. Siempre que intersectamos dos “tidlists”, estamos sólo interesados en la resultante “tidlist” si su cardinalidad alcanza un soporte mínimo, en otro caso debemos eliminar esta intersección. *Eclat* originalmente genera sólo conjuntos de items frecuentes de tamaño superior a 3, aunque se ha modificado para extraer también los conjuntos de items frecuentes de tamaño 1 y 2 [Hipp et al., 2000].

5.2.3 Algoritmo Apriori

Dado el conjunto de transacciones T , estamos interesados en generar todas las reglas que satisfagan ciertas restricciones de dos formas diferentes:

1. Restricciones sintácticas: implican restricciones en los items que pueden aparecer en una regla. Por ejemplo, es posible que estemos interesados sólo en reglas que tienen un item específico I_x que aparezca en el consecuente, o las reglas que tienen un item específico I_y que aparece en el antecedente, o combinaciones de ambas.
2. Restricciones de soporte: están relacionadas con el número de transacciones en T que soporta una regla. Si la regla $X \Rightarrow Y$ tiene una confianza de 1 en la base de datos significa que para el 100% de las transacciones

que contienen ambos atributos, la regla es correcta. Mientras la confianza es una medida de la fuerza de la regla, el soporte se corresponde con la significancia estadística así como con el hecho de que normalmente estamos interesados sólo en las reglas con un soporte por encima de un umbral mínimo dado que si el soporte no es suficientemente grande, significa que no merece la pena tenerla en cuenta o que simplemente puede ser interesante tenerla en cuenta después [Agrawal et al., 1993]. El soporte y la confianza no miden la relación de asociación sino con qué frecuencia los patrones co-ocurren y sólo tienen sentido en la extracción de ARs. El soporte transaccional mínimo (*minsup*) se define como la unión de los items en el consecuente y antecedente de la regla presente en mínimo del $s\%$ de las transacciones en la base de datos. La confianza mínima (*minconf*) simboliza que al menos $c\%$ de las transacciones en la base de datos que satisfacen el antecedente de la regla también satisfacen el consecuente de la regla.

Descubrimiento de ARs

El problema de extraer ARs (“AR discovering” en inglés) es encontrar todas las reglas que satisfacen un soporte y confianza mínimas especificadas por el usuario [Srikant and Agrawal, 1995] y se subdivide en dos subproblemas, en el primero se descubren los conjuntos numerosos de items y en el segundo paso las ARs se derivan de estos conjuntos numerosos de items:

- Encontrar todos los conjuntos de items frecuentes (aquellos cuyo soporte es mayor que el umbral *minsup* especificado por el usuario), también se les denomina conjuntos numerosos de items en la literatura. Este problema puede resolverse construyendo un conjunto candidato de “itemsets” potencialmente frecuentes e identificar los itemsets numerosos dentro de este conjunto candidato. El tamaño de itemsets considerados se incrementa progresivamente hasta que no puedan encontrarse más itemsets frecuentes.
- Generar las ARs derivadas de los conjuntos de items numerosos: Si $X \cup Y$ y X son conjuntos de items frecuentes, la regla $X \implies Y$ se sostiene si la razón entre el soporte($X \cup Y$) y el soporte(X) es al menos, tan grande como el umbral de confianza mínima. Debería tenerse en cuenta que la regla tendrá soporte mínimo dado que $X \cup Y$ es frecuente [Berzal et al., 2001].

Principio Apriori

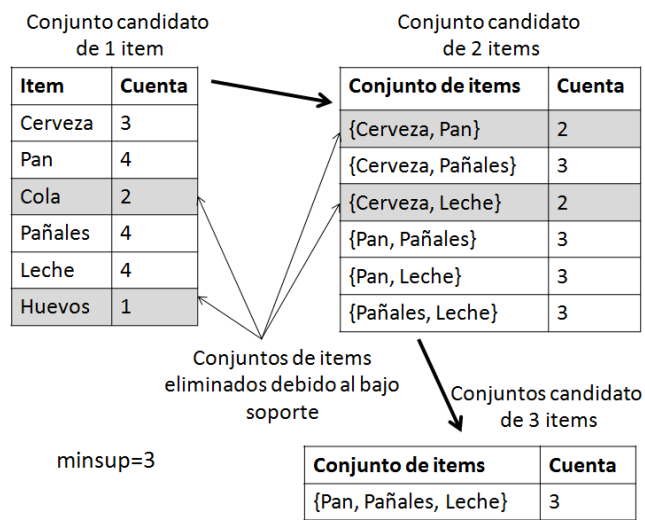


Figura 5.2: Ilustración de la generación de los conjuntos de ítems frecuentes usando el algoritmo Apriori. Fuente [Tan et al., 2005]

Si un conjunto de items es frecuente, todos sus subconjuntos deben serlo también. Del mismo modo, si un conjunto de items es no-frecuente, entonces todos sus subconjuntos lo son y como consecuencia deben ser eliminados [Berzal, 2002].

Existen dos tipos de algoritmos para identificar los conjuntos de items frecuentes o numerosos: secuenciales (que identifican y cuentan los itemsets siguiendo un orden) o paralelos (intentan paralelizar la tarea de encontrar los itemsets). Apriori también se llama *Algoritmo Secuencial*, y es el algoritmo de extracción de ARs más conocido propuesto en [Agrawal and Srikant, 1994]. Los algoritmos para descubrir los conjuntos de items frecuentes hacen múltiples pasadas sobre los datos. En la primera pasada, contamos el soporte de los items individuales y determinamos cuáles de ellos son numerosos, es decir que tengan un soporte mínimo. En cada pasada consecutiva, comenzamos con un conjunto semilla de items que se consideraron numerosos en la pasada anterior. Utilizamos este conjunto semilla para generar itemsets nuevos potencialmente numerosos, llamados conjuntos de items candidatos, y contamos el soporte actual para estos conjuntos de items candidatos mientras que se realiza la pasada sobre los datos. Al final de la pasada, determinamos qué conjuntos de items candidatos son de hecho numerosos, y se convierten en semilla para la siguiente pasada. Este proceso continúa hasta que no se encuentran nuevos conjuntos de items frecuentes [Srikant and Agrawal, 1995]. La notación que utiliza el algoritmo Apriori se puede resumir en:

- L_k : k-itemset con soporte mínimo. Cada miembro de este conjunto tiene dos campos: conjuntos de items y cuenta de soporte.
- C_k : k-itemset candidatos (conjunto de datos potencialmente numerosos). Cada miembro de este conjunto tiene dos campos: conjunto de items y soporte. C_k se genera a partir de L_{k-1} mediante la función *apriorigen*, de esta forma: $C_k = \text{apriorigen}(L_{k-1})$.

La primera pasada del algoritmo descrito en el esquema del algoritmo 1 (“algorithm” en inglés) simplemente cuenta las ocurrencias del item para determinar los conjuntos de 1 item frecuentes. Una sucesiva k -ésima pasada, consta de dos fases. Primero, los itemsets numerosos L_{k-1} encontrados en la pasada $k-1$ se usan para generar los conjuntos de items candidatos C_k , usando la función *apriorigen* que está descrita en el esquema del algoritmo 2. A continuación, la base de datos es escaneada y se cuenta el soporte de los candidatos en C_k . Para contar de forma rápida, necesitamos determinar eficientemente los candidatos en C_k contenidos en una transacción t . Por otro lado, en AprioriTID, se incluye (aparte de L_k y C_k), el término \bar{C}_k : k-itemsets candidatos

Algorithm 1 Algoritmo Apriori

```

1:  $L_1$ =conjuntos numerosos de 1 item;
2: for  $k=2$ ;  $L_{k-1} \neq 0$ ;  $k++$  do
3:    $C_k = \text{apriorigen}(L_{k-1})$ ; //Nuevos candidatos
4:   para todas las transacciones  $t \in D$  empezar
5:    $C_t = \text{subconjunto}(C_k, t)$ ; //Candidatos contenidos en t
6:   for all candidatos  $c \in C_t$  do
7:      $c.cuenta++$ 
8:   end for
9:    $L_k = \{c \in C_k; c.cuenta \geq \text{minsup}\}$ 
10: end for
11: Respuesta:  $\cup_k L_k$ 

```

Algorithm 2 función apriorigen

```

insertar en  $C_k$ 
2: seleccionar  $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-1}$ 
   desde  $L_{k-1p}, L_{k-1q}$ 
4: donde  $p.item_1 = q.item_1, \dots, p.item_{k-2} = q.item_{k-2}, p.item_{k-1} <$ 
    $q.item_{k-1}$ 

```

cuando los TIDs de las transacciones generadas se mantienen asociadas con los candidatos. Se utiliza en AprioriTID para evitar muchos recorridos en bases de datos grandes, utilizando la relación auxiliar en la que se almacena el identificador de cada registro, junto con los itemsets frecuentes que contiene. \bar{C}_k acelera el proceso de extracción de reglas dado que en vez de recorrer todas las transacciones, para ver qué tuplas contienen dichos itemsets, recorreremos C_{k-1} y generamos \bar{C}_k mediante la relación [Berzal et al., 2001] descrita en 5.8:

$$\bar{C}_{k-1}, C_k \longrightarrow \bar{C}_k \quad (5.8)$$

La función *apriorigen* toma como argumento L_{k-1} . Devuelve un conjunto de los k-itemsets, uniendo L_{k-1} con L_{k-1} primero. En la fase de poda, borramos todos los itemsets $c \in C_k$, para eliminar cualquier subconjunto (k-1) de c no esté en L_{k-1} . En la figura 5.2, se proporciona la ilustración de la generación del itemset candidato mediante algoritmo Apriori para las transacciones mostradas en la tabla 5.1. Tal y como Apriori procede, cada item se considera como un conjunto de items candidatos de 1 item, descartando *huevos y cola* porque aparecen en menos de 3 transacciones (*minsup* es 3). La siguiente iteración genera itemsets candidatos de dos items usando sólo los

itemsets frecuentes de 1 ítem dado que el principio Apriori asegura que todos los superconjuntos del conjunto de 1 ítem no frecuente serán no frecuentes también. Y así sucesivamente hasta llegar a la tercera iteración.

Generación de reglas

Una vez obtenidos los itemsets frecuentes, se generan las ARs. Supongamos que los itemsets A_1, B_3, C_2 son frecuentes; también lo serán A_1, B_3 etc. Por lo tanto todas las reglas generadas con el soporte mínimo serían consideradas válidas. Pero en la estrategia de poda para la obtención final de las reglas, entra en juego el umbral de confianza (*minconf*).

Complejidad computacional

La complejidad computacional del algoritmo Apriori se ve afectada por los siguientes factores [Tan et al., 2005]:

1. Umbral de soporte: al bajarlo, se incrementa el tamaño máximo de los itemsets frecuentes y el algoritmo necesita hacer más pasadas sobre los datos.
2. Dimensionalidad (número de ítems): el coste entrada/salida se incrementa debido a un mayor número de itemsets candidatos generados por el algoritmo.
3. Número de transacciones: dado que Apriori hace múltiples pasadas sobre los datos, su tiempo de cómputo se incrementa con un número mayor de transacciones.
4. Promedio de la anchura de la transacción: Para conjuntos de datos densos, la anchura de la transacción promedio puede ser larga. Esto afecta a la complejidad del algoritmo Apriori de dos formas. Primero, el tamaño máximo de los itemsets frecuentes tiende a incrementarse conforme el promedio de la transacción crece y más conjuntos de ítems candidatos deben examinarse. En segundo lugar, conforme el ancho de la transacción se incrementa, más conjuntos de ítems están contenidos en la transacción. Esto incrementará el número de recorridos del árbol hash durante la cuenta de soporte.
5. Generación de conjuntos de 1 ítem frecuentes: para cada transacción, necesitamos actualizar la cuenta de soporte para cada ítem presente en la transacción. Asumiendo que ω es la anchura promedio de la transacción, esta operación requiere un tiempo $O(N\omega)$, donde N es el número total de transacciones.

6. Generación de candidatos: Para generar k-itemsets, pares de (k-1) itemsets se unen para determinar si al menos tienen k-2 items en común. Cada operación de unión necesita al menos (k-2) comparaciones.
7. Cuenta de soporte: el coste computacional para la cuenta de soporte es $O(N \sum_k \binom{n}{k} \alpha_k)$ donde w es la anchura máxima de la transacción y α_k es el coste para actualizar el soporte de la cuenta de k-itemsets candidatos en el árbol hash.

5.2.4 Métodos de discretización

El término continuo se usa en minería de datos para referirse a atributos (características o ROIs que identifican a cada sujeto) que toman un valor numérico (entero o real), o en general a un atributo con un rango ordenado linealmente de valores [Fayyad and Irani, 1993]. Una gran proporción de las bases de datos reales incluyen variables continuas, es decir, medidas en un intervalo o nivel de proporción. Los atributos en un problema de aprendizaje se pueden clasificar en nominales (categóricos) o continuos (numéricos). El proceso de particionar variables continuas en categorías se denomina discretización y su objetivo prioritario es encontrar una serie de puntos de corte para particionar el rango en un número pequeño de intervalos que tengan una buena coherencia de clase. El segundo objetivo consiste en minimizar el número de intervalos sin pérdida significativa de independencia mutua entre los atributos [S. Kotsiantis, 2006]. Los atributos de valor continuo se discretizan previamente a la selección, típicamente particionando el rango de los atributos en subrangos. En general, una discretización es simplemente una condición lógica, en términos de uno o más atributos que sirven para particionar los datos en al menos dos subconjuntos [Fayyad and Irani, 1993].

Generalmente, los métodos de discretización se pueden dividir en: supervisados o no-supervisados, estáticos o dinámicos y locales o globales [Dougherty et al., 1995]. Mientras que los métodos locales producen las particiones que se aplican a regiones localizadas del espacio instancia; los métodos globales, como los intervalos (en inglés “bins”), producen una malla sobre todo el espacio continuo n-dimensional [Chmielewski and Grzymala-Busse, 1994]. En comparación con los no-supervisados, los métodos supervisados hacen uso de las etiquetas de clase en el proceso de discretización. Mientras que los métodos estáticos, como el de intervalo o la partición basada en entropía [Fayyad and Irani, 1993], consiguen en una pasada la discretización de los datos de cada característica y determinan el valor del máximo número de intervalos (K); los

métodos dinámicos llevan a cabo una búsqueda para todas las características simultáneamente con el objeto de capturar las interdependencias en la discretización de características. El método de “bin” de igual anchura (el utilizado en esta tesis para el análisis de imagen funcional, véase capítulo 6) es no supervisado y a menudo se aplica como una media para producir valores nominales desde los continuos involucrando dividir el rango de los valores observados de una característica continua en K “bins” (siendo esta cifra seleccionada por el usuario) de igual tamaño [Dougherty et al., 1995]. Si se observa que una variable x tiene valores limitados por x_{min} y x_{max} , entonces este método computa el ancho del “bin” como:

$$\alpha = \frac{x_{max} - x_{min}}{K} \quad (5.9)$$

y construye los límites del “bin” en $x_{min} + i \cdot \alpha$ donde $i=1, \dots, K-1$.

Discretización y ARs

La extracción de ARs sobre atributos discretizados, se propone en [Ribeiro et al., 2009], tras la extracción de características de las imágenes de entrenamiento, se propone el algoritmo Omega que lleva a cabo la discretización de valores continuos. Omega procesa cada característica de forma separada y discretiza un rango de N valores ordenados en $4N$ pasos, teniendo un coste lineal de N . Tras el algoritmo Omega, se lleva a cabo la extracción de ARs que se utilizan como entrada para el algoritmo ACE, que utiliza las ARs para sugerir palabras clave y de esta forma diagnosticar la imagen de test.

En [Park and Lee, 2009], se presenta un método de discretización basado en máquinas de vectores de soporte (SVM). Mediante el hiperplano $h : \mathbf{a} \cdot \mathbf{x} + b = 0$ encontrado por SVM, los atributos continuos se transforman en atributos categóricos cuyo rango consiste en dos valores tales como:

$$x^s \rightarrow z^h = \begin{cases} h^+ & \text{si } a \cdot x^s + b > 0 \\ h^- & \text{si } a \cdot x^s + b < 0 \end{cases} \quad (5.10)$$

Otra forma de discretización consiste en combinar SVM y la discretización basada en entropía; en lugar de utilizar los signos de $a \cdot x^s + b$, los atributos continuos se transforman en un atributo continuo de una dimensión tal como: $x^s \rightarrow a \cdot x^s + b$. El valor $a \cdot x^s + b$ es la distancia con signo desde un punto x^s al hiperplano separador. Entonces, la discretización basada en entropía se puede conseguir en el espacio de características transformado. De todas las particiones binarias posibles, se elige la que minimiza la entropía total y se aplica recursivamente para los subintervalos divididos hasta que se encuentre un criterio de parada [Fayyad and Irani, 1993]. A partir de los datos discretizados,

se utilizan clasificadores asociativos, que combinan las reglas de asociación con la clasificación. En [Liu et al., 1998a], se propone la discretización de atributos continuos para integrar clasificación y extracción de ARs. En el capítulo 6 se describe la extracción de ARs sobre atributos discretos para el análisis de imagen funcional.

5.3 Métodos de extracción de características

Los métodos de extracción de características (FE, del inglés “Feature Extraction”) pretenden la reducción de la dimensionalidad posterior a la FS para la elección de las características más adecuadas en clasificación [Güçlü et al., 2011], [Huang and Xiao, 2011], [Liu et al., 2012]. Algunos de los métodos de FE más comunes son: ICA, análisis discriminante lineal (LDA, del inglés “Linear Discriminant Analysis”), PCA, PLS etc. ICA busca aquellas direcciones en el espacio de características que muestran la independencia de señales, siendo particularmente útil para segmentar las señales de múltiples fuentes [Duda et al., 2001]. En LDA [Sharma and Paliwal, 2008], intentamos maximizar la dispersión entre clases con respecto a la dispersión dentro de la clase. Este proceso requiere la resolución del problema de descomposición de eigenvalores, que implica el cálculo de la matriz de dispersión entre clases. En este trabajo se utilizan PCA y PLS como extractores de características (posterior a la selección previa con ARs).

5.3.1 PCA

PCA es una poderosa técnica de extracción de características que lleva a cabo la proyección de los datos originales de dimensión D en un subespacio de menor dimensión mediante una transformación ortogonal que minimiza el error cuadrático medio [Jolliffe, 2002]. PCA pretende encontrar las componentes principales (CPs) o autovectores de la matriz de covarianza de un conjunto de imágenes tratadas cada una como un punto (o vector) en un espacio de alta dimensión. Los autovectores quedan ordenados en función de la cantidad de variación entre las imágenes que representan [Pérez, 2010]. En primer lugar, dados un conjunto de vectores centrados \mathbf{x}_k con $k = 1, \dots, M$ que verifican $\mathbf{x}_k \in \mathbb{R}^N$ y $\sum_{k=1}^M \mathbf{x}_k = 0$ definimos la matriz de covarianza que PCA diagonaliza como [B. Scholkopf, 1998]:

$$C = \frac{1}{M} \cdot \sum_{j=1}^M \mathbf{x}_j \cdot \mathbf{x}_j^T \quad (5.11)$$

Para la diagonalización de PCA, se tiene que resolver la ecuación de eigenvalor:

$$\lambda \cdot \mathbf{v} = C \cdot \mathbf{v} \quad (5.12)$$

para eigenvalores $\lambda \geq 0$ y $\mathbf{v} \in \mathbb{R}^N - 0$. Dado que:

$$C \cdot v = \frac{1}{M} \cdot \sum_{j=1}^M (\mathbf{x}_j \cdot \mathbf{v}) \cdot \mathbf{x}_j^T \quad (5.13)$$

todas las soluciones \mathbf{v} con $\lambda \neq 0$ deben caer en el alcance de \mathbf{x}_k con $k = 1, \dots, M$, por lo tanto la ecuación anterior es equivalente a:

$$\lambda \cdot (\mathbf{x}_k \cdot \mathbf{v}) = \mathbf{x}_k \cdot C \cdot \mathbf{v} \quad (5.14)$$

donde $\mathbf{x}_k \cdot \mathbf{v}$ son los autovectores de la matriz de covarianza. Concluimos que los autovalores nos permiten ordenar sus autovectores asociados de acuerdo a la utilidad de los mismos a la hora de caracterizar la variación entre imágenes [Pérez, 2010].

Kernel PCA

Kernel PCA es una generalización no lineal de PCA [B. Scholkopf, 1998] en la que cada vector \mathbf{x} se proyecta desde el espacio de entrada, \mathbb{R}^n , a un espacio de características de alta dimensión \mathbb{R}^f mediante una función de mapeado no-lineal donde $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^f$ con $f > n$. Definimos la matriz de covarianza como:

$$C^\phi = \frac{1}{M} \cdot \sum_{j=1}^M \phi(\mathbf{x}_j) \cdot \phi(\mathbf{x}_j)^T \quad (5.15)$$

En \mathbb{R}^f , para la diagonalización de la matriz de covarianza, el problema correspondiente de eigenvalor es:

$$C^\phi \cdot \omega^\phi = \lambda \cdot \omega^\phi \quad (5.16)$$

Todas las soluciones ω^ϕ con $\lambda \neq 0$ están en el espacio $\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)$ donde N es el número de muestras, y existen coeficientes α_i tales que:

$$\omega^\phi = \sum_{i=1}^N \alpha_i \cdot \phi(\mathbf{x}_i) \quad (5.17)$$

denotando una matriz kernel de $N \times N$ por:

$$K_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \quad (5.18)$$

el problema del kernel PCA se transforma en:

$$N \cdot \lambda \cdot K \cdot \alpha = K^2 \cdot \alpha = N \cdot \lambda \cdot \alpha = K \cdot \alpha \quad (5.19)$$

donde α denota un vector columna con entradas $\alpha_1, \dots, \alpha_N$. La derivación anterior asume que todas las muestras proyectadas $\phi(\mathbf{x})$ se centran en \mathbb{R}^f . Además existen dos ventajas fundamentales del kernel PCA no lineal; en primer lugar, las componentes principales no lineales logran mejores ratios de reconocimiento, y en segundo lugar, el rendimiento para las componentes no lineales pueden mejorarse usando más componentes de las posibles en el caso lineal [López et al., 2011].

5.3.2 PLS

PLS es una importante técnica de extracción de características que crea vectores conocidos como latentes y maximiza la covarianza entre diferentes conjuntos de variables [Gerlach et al., 1979], [F.Segovia, 2010]. En particular, esta maximización se lleva a cabo entre el peso de las variables observadas (\mathbf{X}) y las etiquetas (\mathbf{Y}) mediante variables lineales latentes de acuerdo a modelos de regresión [Bastien et al., 2005]. Para ello se utiliza el algoritmo SIMPLS propuesto por Sijmen de Jong en 1993 [de Jong, 1993] como alternativa al algoritmo NIPALS para PLS. La principal diferencia con NIPALS es el grado de deflación ya que en SIMPLS, no se hace deflación de las matrices de datos centradas \mathbf{X} e \mathbf{Y} pero esta se lleva a cabo para la matriz de covarianza, o con más precisión, para la matriz de producto cruzado $S = \mathbf{X}^T \cdot \mathbf{Y}$ [Varmuza and Filzmoser, 2009]. Por otro lado, el método de kernel PLS no lineal se basa en el mapeado de los datos de entrada originales en un espacio de características de alta dimensión [Rosipal, 2003].

La principal diferencia entre PLS y PCA radica en que el primero crea vectores ortogonales de pesado maximizando la covarianza entre \mathbf{X} e \mathbf{Y} , es decir, PLS no sólo considera la varianza de las muestras sino que también tiene en cuenta la etiqueta de clase [Kettaneh et al., 2005]. En esta tesis se propone el algoritmo LMNN como técnica de FE para el análisis de imágenes funcionales (véase capítulo 7) y se comenta teóricamente en la siguiente sección 5.4 dedicada a LMNN.

5.4 LMNN

La métrica basada en distancia [Yang and Jin, 2006] es un asunto clave en muchos algoritmos de máquinas de aprendizaje y en los últimos años, se ha utilizado el principio de amplio margen con asiduidad. De hecho, las funciones de pérdidas basadas en margen, en lugar de margen en sí mismo, se utilizan para caracterizar la complejidad de la clasificación. AdaBoost fue publicado en [Breiman, 1999] para minimizar el margen de funciones de pérdida exponencial. Mason propuso procedimientos de clasificación AnyBoost que llevan a cabo gradiente descendente en el espacio de funciones con funciones de pérdida general del margen [Mason et al., 2000]. Algunas técnicas fueron desarrolladas en [Hu et al., 2011] para el aprendizaje del peso de las características en la clasificación de los vecinos más cercanos.

En este trabajo se propone LMNN para la clasificación de los k vecinos más cercanos (KNN) desde los ejemplos etiquetados usando *la función de pérdidas* [Weinberger et al., 2009]. KNN es un algoritmo no lineal para clasificación bastante conocido y que trabaja típicamente con la distancia Euclídea y de manera adecuada si el conjunto de entrenamiento no es demasiado extenso. Cada elemento x se compara a través de una función de distancia $dist(\mathbf{x}, \mathbf{x}_i)$ con todos los puntos \mathbf{x}_i ($i = 1, \dots, n$) del conjunto de entrenamiento. Las k distancias mínimas se calculan y la mayoría de las etiquetas \mathbf{y}_i de ellas se toman como la etiqueta resultante para \mathbf{x} [S. Lemm, 2011]. En resumen LMNN se usa en este trabajo de diferentes formas: *i*) como una transformación del espacio de características obtenida por medio de PLS o PCA para separar mejor las clases control y AD, *ii*) como técnica de reducción de características actuando la transformación como una matriz rectangular (LMNN-RECT) y *iii*) como clasificador.

El objetivo de LMNN es obtener una familia de métricas sobre el espacio de características. El conjunto de entrenamiento se denota por $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ de n ejemplos etiquetados con entradas $\mathbf{x}_i \in \mathbb{R}^d$ y etiquetas de clase asociadas \mathbf{y}_i . Nuestro objetivo es aprender una transformación lineal $\mathbf{L}: \mathbb{R}^d \rightarrow \mathbb{R}^d$. Estas métricas calculan distancias cuadradas como:

$$\mathbf{D}_{\mathbf{L}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2^2 \quad (5.20)$$

La ecuación 5.20 se usa comúnmente para expresar distancias cuadradas en términos de la matriz \mathbf{M} :

$$\mathbf{M} = \mathbf{L}^T \cdot \mathbf{L} \quad (5.21)$$

Por otro lado, las distancias cuadradas se denotan mediante métricas de Ma-

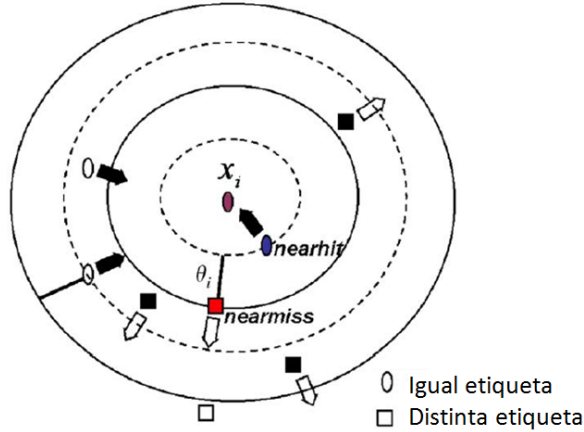


Figura 5.3: Ilustración para la optimización de la función de pérdidas. Obtenida de [Li and Lu, 2009].

halanobis en términos de \mathbf{M} :

$$\mathbf{D}_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \cdot \mathbf{M} \cdot (\mathbf{x}_i - \mathbf{x}_j) \quad (5.22)$$

Una distancia de Mahalanobis puede estar parametrizada en términos de la matriz \mathbf{L} o de la matriz \mathbf{M} [Weinberger et al., 2009]. La primera no tiene restricciones, mientras que la segunda debe ser semidefinida positiva. La idea fundamental de LMNN consiste en minimizar la función de pérdidas (véase la subsección siguiente) que es capaz de aprender una métrica de distancia bajo la cual las entradas y los vecinos objetivo estén más unidos.

5.4.1 Función de pérdidas

La función de pérdidas es una técnica comúnmente utilizada en el aprendizaje automatizado para encontrar el balance correcto entre margen de error pequeño y amplio margen [Li and Lu, 2009]. Una vez que se elige la función de pérdidas, el objetivo del algoritmo de aprendizaje es minimizarlo y obtener un margen amplio.

En LMNN, se pretende que los vecinos objetivo, definidos como patrones de entrada de la misma clase, estén más juntos. La función de pérdidas a minimizar consiste en dos términos. Uno actúa para favorecer la *unión* de los vecinos objetivo penalizando las largas distancias entre cada entrada y sus vecinos objetivo. El otro término actúa para presionar la *separación* de ejemplos etiquetados de forma diferente. Esto penaliza las pequeñas distancias

entre ejemplos etiquetados de forma diferente [Chaves et al., 2012d]. En este segundo término, especialmente nos preocupamos de las muestras etiquetadas de forma diferente localizadas a una distancia desde x_i a cualquiera de sus vecinos objetivo más un margen determinado, que es una constante positiva definida como en la ecuación 5.23:

$$\theta_i = \left| \|\mathbf{x}_i - \text{nearmiss}(\mathbf{x}_i)\|^2 - \|\mathbf{x}_i - \text{nearhit}(\mathbf{x}_i)\|^2 \right| \quad (5.23)$$

donde (como se muestra en la figura 5.3) $\text{nearhit}(x_i)$ y $\text{nearmiss}(x_i)$ denotan las muestras más cercanas a x_i con la misma y distinta etiqueta respectivamente. El término que favorece la separación amplia entre ejemplos etiquetados de forma diferente está representado por la siguiente ecuación:

$$\varepsilon_{\text{separar}}(\mathbf{L}) = \sum_{j \rightarrow i} \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 \quad (5.24)$$

donde $j \rightarrow i$ significa que la entrada \mathbf{x}_j es un vecino objetivo de la entrada \mathbf{x}_i . Una nueva variable indicador se introduce para definir el término que une en la función de pérdidas:

$$\mathbf{y}_{il} = \begin{cases} 1 & \text{if } \mathbf{y}_i = \mathbf{y}_l \\ 0 & \text{otro caso} \end{cases} \quad (5.25)$$

de tal forma que:

$$\varepsilon_{\text{unir}}(\mathbf{L}) = \sum_{i,j \rightarrow i} \sum_l (1 - \mathbf{y}_{il}) \cdot [1 + \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 - \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|^2]_+ \quad (5.26)$$

donde $[\]_+ = \max(z, 0)$ denota la función de pérdidas estándar que se define en [Weinberger et al., 2009]. Finalmente, combinamos los dos términos $\varepsilon_{\text{separar}}(\mathbf{L})$ y $\varepsilon_{\text{juntar}}(\mathbf{L})$ en una sola función de pérdidas para aprendizaje de la métrica de distancias. Los dos términos pueden tener efectos competitivos, atraer vecinos cercanos y repeler impostores. Los impostores se definen como entradas con diferentes etiquetas. Un parámetro de pesado $\mu \in [0, 1]$ balancea estos objetivos.

$$\varepsilon(\mathbf{L}) = (1 - \mu) \cdot \varepsilon_{\text{separar}}(\mathbf{L}) + \mu \cdot \varepsilon_{\text{unir}}(\mathbf{L}) \quad (5.27)$$

La figura 5.3 muestra el comportamiento de la función de pérdidas para $k = 3$ vecinos objetivo. Se demuestra que los vecinos objetivo (círculos) se mueven cerca de x_i , mientras que los impostores (cuadrados) se alejan de x_i consiguiendo mejoras en la clasificación.

5.4.2 Transformación LMNN y LMNN-RECT

El objetivo del aprendizaje de métricas de distancias puede ser formulado de dos formas: aprender una transformación lineal L sobre los datos de entrada: $\mathbf{x}_i \rightarrow L \cdot \mathbf{x}_i$ o aprender una métrica de Mahalanobis $M = L \cdot L^T$.

La función de pérdidas [Weinberger et al., 2009] necesita ser optimizada para obtener la métrica de distancia que transforma en términos de la matriz de transformación de bajo rango, L . La transformación sobre L no es convexa al contrario que la optimización sobre M , pero se puede calcular computacionalmente un mínimo posiblemente local por los métodos basados en gradientes estándar. Si L es una matriz cuadrada (transformación LMNN), la transformación no reduce la dimensionalidad, pero esta solución sirve para rotar y reordenar las coordenadas de la entrada por sus respectivas varianzas. Si L es una matriz rectangular (LMNN-RECT), la transformación lineal proyecta las entradas en un subespacio de dimensionalidad menor. Llamamos a esta aproximación LMNN-RECT [Weinberger and Saul, 2008], en la que L es una matriz con un tamaño igual al número de características seleccionadas en una dimensión, mientras que la otra dimensión es el número de características que pretende reducirse.

5.4.3 Kernel LMNN

Resulta interesante considerar el caso en el que \mathbf{x}_i se mapea en un espacio de características de alta dimensión $\phi(\mathbf{x}_i)$ y se busca una distancia de Mahalanobis en este espacio. Nos centramos en el caso en el que los productos de puntos en el espacio de características puede expresarse vía una función de kernel tal que:

$$\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j) \quad (5.28)$$

para algún kernel k [Globerson and Roweis, 2005]. Cuando usamos el marco del método *atajo* de Kernel PCA, el método original LMNN puede ser inmediatamente utilizado como Kernel LMNN (KLMNN) como se explica en [Chatpatanasiri et al., 2010]. Este nuevo marco ofrece varias ventajas prácticas sobre el conjunto clásico de kernel, ya que no hay fórmulas matemáticas y no se requiere reprogramación para una implementación de kernel, siendo ésta una forma de acelerar el algoritmo sin trabajo extra, y evitando también algunos problemas como la singularidad.

5.5 Clasificadores

Un clasificador es una función que mapea una instancia sin etiqueta a una etiqueta utilizando estructuras de datos internas. Sea V el espacio de instancias sin etiqueta e Y el conjunto posible de etiquetas. Sea $\chi = V \times Y$ el espacio de instancias etiquetadas y $D = \{x_1, x_2, \dots, x_n\}$ un conjunto de datos (posiblemente multiconjunto), que consiste en n instancias etiquetadas, donde $\langle v_i \in V, y_i \in Y \rangle$. Un clasificador C mapea una instancia sin etiqueta $v_i \in V$ a una etiqueta $y_i \in Y$. Algo que se ha tenido en cuenta en este trabajo, es la capacidad de generalización de un clasificador, una virtud que favorece la capacidad de operar con nuevos datos independientes, es decir la capacidad de obtener buenos resultados para diferentes bases de datos. A continuación describimos algunos métodos de clasificación.

5.5.1 Clasificación basada en técnicas de extracción de ARs

La clasificación y el descubrimiento de las ARs son similares excepto en que la clasificación involucra la predicción de un atributo, es decir la clase, mientras que el descubrimiento de las ARs puede predecir cualquier atributo en el conjunto de datos [Nofal and Ahmad, 2010]. En los últimos años, ha surgido una nueva aproximación que integra la extracción de ARs con clasificación. Algunos clasificadores exactos y efectivos basados en la clasificación asociativa se han presentado recientemente, tales como: CPAR [Yin and Han, 2003], CMAR [Li et al., 2001], MMAC [Thabtah et al., 2004] y CBA [Liu et al., 1998b]. Muchos estudios experimentales mostraron que la clasificación basada en extracción de ARs es una aproximación de gran potencial que construye más sistemas de clasificación exactos y predictivos que los métodos tradicionales de clasificación como por ejemplo los árboles de decisión.

Las técnicas de clasificación asociativas generan items frecuentes haciendo más de un escáner en el conjunto de entrenamiento. En el primer escáner, encuentran el soporte de los 1-itemsets que se utilizan para encontrar los 2-itemsets frecuentes que son la entrada para encontrar los 3 itemsets frecuentes y así sucesivamente. Cuando se han descubierto los items frecuentes, la clasificación basada en algoritmos de ARs extrae un conjunto completo de ARs de clase (CAR, del inglés “Classification Association Rules”) para aquellos items frecuentes que superan *minconf*.

Uno de los primeros algoritmos para unir clasificación con ARs se propuso en [Liu et al., 1998b]. Esta aproximación de clasificación consiste en dos fases principalmente, la fase 1 implementa el algoritmo Apriori para descubrir items frecuentes. La fase 2 involucra la construcción del clasificador. Se define T

como el conjunto de datos de entrenamiento con m atributos y $|T|$ filas. P es una lista de etiquetas de clase. Un item se define como la asociación de un atributo y su valor (AT_i, a_i) o una combinación entre 1 y m valores de atributos diferentes. Una regla r para clasificación se representa en la forma:

$$(AT_{i1} = x_{i1}) \wedge (AT_{i2} = x_{i1}) \wedge \dots \wedge (AT_{in} = x_{in}) \rightarrow p_{i1} \quad (5.29)$$

donde el antecedente de la regla es un item y el consecuente es una clase. La apariencia ($Appr$) de la regla r en T es el número de veces que el antecedente de la regla ha aparecido en T . La frecuencia de soporte ($SuppFreq$) de r es el número de casos en T que relacionan el antecedente de r y pertenecen a la clase p_i . Una regla r supera el umbral mínimo de soporte ($minsup$) si para r :

$$SuppFreq(r)/|T| \geq minsup \quad (5.30)$$

Una regla r pasa el umbral de confianza mínimo ($minconf$) si:

$$SuppFreq(r)/Appr(r) \geq minconf \quad (5.31)$$

Cualquier item en T que supere el $minsup$ se considera item frecuente.

El problema de producir reglas con múltiples etiquetas se ha investigado recientemente con la clasificación asociativa multiclase (MMAC). Los primeros algoritmos que motivaron la idea de utilizar una AR para clasificación fueron los CBA [Liu et al., 1998b]. CBA implementa el algoritmo Apriori que requiere múltiples pasadas sobre el conjunto de entrenamiento para descubrir items frecuentes. Una vez que el descubrimiento de items frecuentes terminó en la fase 1; en la 2, CBA lleva a cabo la conversión de cualquiera de ellos que superen en una regla un $minconf$.

Clasificación basada en evaluación de reglas

De una forma similar a los clasificadores asociativos, la clasificación basada en evaluación de ARs en este trabajo se implementa en dos fases: fase de entrenamiento (extracción de reglas) y fase de test (clasificación). Las reglas extraídas presentan asociaciones en el grupo control, dado que este presenta un patrón de hipoperfusión continuo y típico [Nitrini et al., 2000] para la mayoría de los sujetos de la base de datos pero no para todos los pacientes con AD. En este sentido la clasificación presupone que los normales van a verificar un número superior de ARs que los AD.

- Fase de entrenamiento: se lleva a cabo la extracción de ARs. Las ARs asocian las ROIs de los controles usando la estrategia de validación

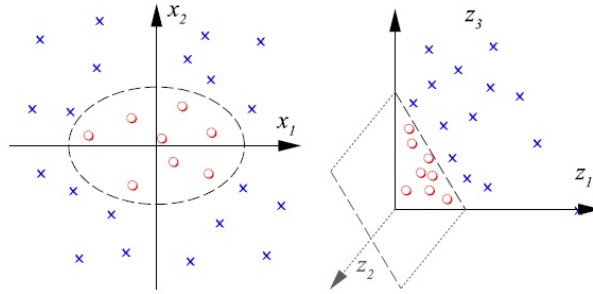


Figura 5.4: Ejemplo de clasificación bidimensional. A la derecha, las dos clases se pueden separar en el espacio de características por un hiperplano lineal. A la izquierda, la separación entre clases se realiza mediante una frontera de decisión elipsoidal [Müller et al., 2001].

cruzada [Gidskehaug et al., 2008], [Stone, 1974]. Cuando un control específico se está evaluando en la fase de test, no se considera en la extracción de reglas, evitando el overtraining. Si la disponibilidad de las muestras es limitada, la separación de los datos en entrenamiento y validación puede decrementar la calidad de tanto el modelo de calibración como en el de la validación. Por esta razón, esto es resuelto procesando toda la información disponible (scans), para estimar el error residual del modelo en la fase de test [Gidskehaug et al., 2008] en un proceso iterativo LOO.

- Fase de test: En esta fase se comprueba el número de ARs que son verificadas por cada sujeto de test siguiendo la estrategia de validación LOO. Si un sujeto verifica un número mayor o igual de ARs que un *umbral de clase* entonces se clasifica como CTRL, en otro caso como AD.

Este modelo de clasificación se aplica en el capítulo 6 para ARs aplicadas sobre atributos continuos, discretos o discriminantes.

5.5.2 SVM

Las SVMs se hicieron populares hace algunos años para resolver problemas de clasificación, regresión y detección. Una propiedad importante de SVM

es que la determinación de los parámetros del modelo se corresponde con un problema de optimización convexa, y de esta forma cualquier solución local es también un óptimo global. En SVM la superficie límite se define como única para la maximización del margen de separación entre clases. El margen se define como la distancia perpendicular entre la frontera de división y los puntos de los datos más cercanos. El hecho de maximizar el margen conlleva la elección particular de la superficie de decisión. La localización de la frontera se determina por un subconjunto de puntos de datos conocidos como vectores de soporte que indican el margen de separación entre clases. La discusión de SVM comienza con el problema de clasificación de dos clases usando los modelos lineales de la forma [Bishop, 2006]:

$$y(\mathbf{x}) = \mathbf{w}^T \cdot \phi(\mathbf{x}) + b \quad (5.32)$$

donde $\phi(x)$ denota la transformación del espacio de características fijado, y b el parámetro de desplazamiento explícito. El conjunto de entrenamiento comprende N vectores de entrada $\mathbf{x}_1, \dots, \mathbf{x}_N$ con los correspondientes valores objetivo t_1, \dots, t_N donde $t_n \in \{-1, 1\}$ y los datos \mathbf{x} se clasifican de acuerdo al signo de $y(\mathbf{x})$. Si se asume que el conjunto de entrenamiento es linealmente separable en el espacio de características, se satisface que $y(\mathbf{x}_n) > 0$ para puntos con $t_n = +1$ e $y(\mathbf{x}_n) < 0$ para puntos con $t_n = -1$, de tal forma que $t_n \cdot y(\mathbf{x}_n) > 0$ para todos los datos de entrenamiento. La solución de margen máximo se obtiene resolviendo:

$$\operatorname{argmax}_{\mathbf{w}, b} = \left\{ \frac{1}{\|\mathbf{w}\|} \cdot \min_n [t_n \cdot (\mathbf{w}^T \cdot \phi(\mathbf{x}_n)) + b] \right\} \quad (5.33)$$

Mediante un reescalado, convertimos el problema en equivalente:

$t_n \cdot (w^T \cdot \phi(x_n) + b) = 1$ con $n = 1, \dots, N$ para el punto que esté más cercano a la superficie de separación. En este caso, todos los puntos de los datos satisfacen las restricciones: $t_n \cdot (w^T \cdot \phi(x_n) + b) \geq 1$ con $n = 1, \dots, N$. El problema de optimización requiere simplemente que maximicemos $\|w\|^{-1}$ lo que requiere resolver el problema de optimización $\operatorname{argmin}_{\mathbf{w}, b} \frac{1}{2} \cdot \|\mathbf{w}\|^2$. Este necesita las restricciones impuestas por los multiplicadores de Lagrange dada la función de Lagrange $L(\mathbf{w}, b, \mathbf{a})$ (que introduce multiplicadores de Lagrange $a_n \geq 0$) obteniendo la representación dual del problema de amplio margen en el cual maximizamos:

$$\tilde{L}(a) = \sum_{i=1}^N a_i - \frac{1}{2} \cdot \sum_{n=1}^N \sum_{m=1}^N a_n \cdot a_m \cdot t_n \cdot t_m \cdot k(\mathbf{x}_n, \mathbf{x}_m) \quad (5.34)$$

que está sujeto a las restricciones $a_n \geq 0$ con $n = 1, \dots, N$ y $\sum_{n=1}^N a_n \cdot t_n = 0$. La función de kernel (especialmente útil para problemas en los que la separación es no lineal) está definida como: $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \cdot \phi(\mathbf{x}')$. Para clasificar puntos nuevos usando el modelo entrenado, evaluamos el signo de $y(\mathbf{x})$ que puede ser expresado en términos de los parámetros $\{a_n\}$ y sustituyendo \mathbf{w} por la función de kernel dando lugar a:

$$y(x) = \sum_{i=1}^N a_n \cdot t_n \cdot k(\mathbf{x}, \mathbf{x}_n) + b \quad (5.35)$$

que está sujeto a las condiciones de Karush-Kuhn-Tucker (KKT) [Wu, 2009].

El espacio de entrada no necesita ser explícitamente mapeado a un espacio de características de gran dimensión en términos de una función no lineal: $\phi : x \rightarrow \phi(x)$ [Vapnik, 1995]. En lugar de eso, los métodos basados en kernel [Salas-Gonzalez, 2009] se aprovechan del hecho de que la mayoría de los métodos lineales sólo requieren el cómputo de los productos por puntos. Por consiguiente, el truco del aprendizaje basado en kernel consiste en sustituir una función de kernel apropiada en el espacio de entrada para los productos por puntos en el espacio de características, es decir:

$$k : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R} \quad (5.36)$$

del tal forma que: $k(x, x') = \phi(x)^T \cdot \phi(x')$. Los kernels más comúnmente utilizados son [S. Lemm, 2011]:

$$\text{Gaussiano} : k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2 \cdot \sigma^2}\right), \sigma > 0 \quad (5.37)$$

$$\text{Polinomial} : k(x, x') = (x^T \cdot x' + c)^d \quad (5.38)$$

$$\text{Sigmoide} : k(x, x') = \tanh(\kappa(x^T \cdot x') + \theta), \kappa > 0, \theta < 0 \quad (5.39)$$

Hasta el momento hemos supuesto que los datos de entrenamiento puede separarse de forma lineal o no-lineal pero siendo esto un caso idílico dado que en la práctica se puede solapar las distribuciones de clase condicional, en los que la separación exacta de los datos de entrenamiento puede implicar una generalización pobre. Para resolver el problema de optimización cuando las clases no son completamente separables, se introducen variables de pérdidas $\xi_n \geq 0$ donde $n = 1, \dots, N$ con una variable de pérdidas para cada punto de

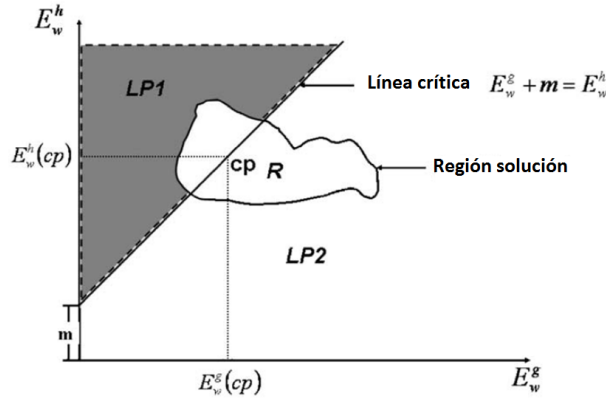


Figura 5.5: El plano formado por el modelo de energía. Obtenido de [Li and Lu, 2009]

entrenamiento. Se definen como $\xi_n = 0$ para datos dentro del margen correcto y $\xi_n = |t_n - y(\mathbf{x}_n)|$ para los otros puntos.

Las restricciones de clasificación se reemplazan por: $t_n \cdot y(\mathbf{x}_n) \geq 1 - \xi_n$ con $n = 1, \dots, N$ en las que las variables de pérdidas se restringen a satisfacer $\xi_n \geq 0$. Los puntos con $0 < \xi_n \leq 1$ caen dentro del margen, pero en el lado correcto, y aquellos puntos para los que $\xi_n \geq 1$ se clasifican mal. Nuestro objetivo es ahora maximizar el margen y penalizar suavemente los puntos que caen en el lugar incorrecto de la frontera. Por lo tanto hay que minimizar:

$$C \cdot \sum_{i=1}^N \xi_n + \frac{1}{2} \cdot \|\mathbf{w}\|^2 \quad (5.40)$$

Adicionalmente y aunque no se aplica experimentalmente en este trabajo dado que clasificamos únicamente en dos clases, SVM puede aplicarse en problemas multiclase, por ejemplo construyendo K SVMs separados o bien entrenando conjuntamente los K SVMs basándose en maximizar el margen de cada una respecto a las clases restantes entre otras posibilidades [Bishop, 2006]. SVM se usó en combinación con los atributos seleccionados con ARs y extraídos con PCA o PLS, como se explica en el capítulo 6.

5.5.3 LMNN

Algunas técnicas se han desarrollado para aprender a pesar la importancia de las características con el objetivo de controlar la variación de la distancia en la clasificación de vecinos cercanos. La distancia Euclídea (la más

común), asume que cada característica es igual de importante e independiente que otras. Como contraste, una métrica de distancia de calidad como Mahalanobis, debe identificar características relevantes asignando diferentes pesos o factores importantes a las ROIs extraídas [Xiang et al., 2008]. Sólo cuando las características no están correladas, la distancia de Mahalanobis es idéntica a la Euclídea. Por otro lado, nuestro trabajo se ha motivado por el aprendizaje de la métrica basada en energía (EBC), obteniendo con ella los mejores resultados [Chopra et al., 2005]. Siendo \mathbf{x}_i y \mathbf{x}_j un par de muestras, w es el vector compartido de parámetros, la energía entre muestras se define como [Li and Lu, 2009]:

$$E(w, \mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_w \quad (5.41)$$

que se utiliza para medir la compatibilidad entre \mathbf{x}_i y \mathbf{x}_j , mediante una distancia euclídea pesada. Dado un par genuino de muestras ($\mathbf{x}_i, \mathbf{x}_j$) en el que x_j es un vecino objetivo de \mathbf{x}_i , y un par de muestras heterogéneo ($\mathbf{x}_i, \mathbf{x}_p$) en el que x_p es una muestra etiquetada diferente cuya distancia a \mathbf{x}_i no excede la distancia desde \mathbf{x}_i a cualquiera de sus vecinos objetivo más el de un margen θ_i (descrito en la sección 5.4.1 de la función de pérdidas de este capítulo), es decir:

$$\exists \theta_i > 0, E(w, \mathbf{x}_i, \mathbf{x}_j) + \theta_i < E(w, \mathbf{x}_i, \mathbf{x}_p) \quad (5.42)$$

siendo L la función de pérdidas total que consiste en la suma de las funciones de pérdidas parciales para un par genuino y heterogéneo (L_g y L_h) respectivamente. Como se muestra en la figura 5.5, existen dos planos $E(w, \mathbf{x}_i, \mathbf{x}_j) + \theta_i < E(w, \mathbf{x}_i, \mathbf{x}_p)$ y $E(w, \mathbf{x}_i, \mathbf{x}_j) + \theta_i \geq E(w, \mathbf{x}_i, \mathbf{x}_p)$, que se denotan LP1 y LP2 respectivamente. La idea es minimizar L sobre $E(w, \mathbf{x}_i, \mathbf{x}_j)$ (también llamado E_g^w) y $E(w, x_i, x_p)$ (denominado E_h^w) para todos los valores de w en su dominio para que se cumpla la condición de la fórmula 5.42, siendo la región solución R y el punto óptimo (cp) (que están representadas en la figura 5.5).

EBC consiste en calcular la función de pérdidas para cada etiqueta posible \mathbf{y}_i . Para ello se lleva a cabo la minimización de tres términos. El primero se define como las distancias cuadradas a los k vecinos objetivo de \mathbf{x}_i . El segundo término acumula la función de pérdidas sobre todos los impostores (etiquetados diferente) que invaden el perímetro alrededor de \mathbf{x}_i determinado por sus vecinos objetivo. El tercer término es la acumulación de la función de pérdidas para los ejemplos etiquetados diferente cuyos perímetros están invadidos por \mathbf{x}_i [Weinberger et al., 2009].

5.5.4 Comparativa LMNN y SVM

Los clasificadores LMNN y SVM se usan en este trabajo para construir sistemas CAD para la AD (véase capítulo 7 para más detalles). Estos presentan muchas similitudes, por ejemplo su potencial para trabajar con espacios de características no lineales mediante el uso del atajo del kernel (“kernel trick”) o el hecho de que utilizan una función objetivo convexa para lograr la maximización del margen de separación entre clases diferentes [Weinberger et al., 2009]. LMNN puede considerarse como un homólogo de SVM reemplazando la clasificación kNN por la clasificación lineal de SVM. Sin embargo, LMNN se diferencia de SVM en que no requiere modificación para los problemas multiclase los cuales implican la combinación de resultados aportados por muchos clasificadores binarios, ya que no existe una dependencia explícita en el número de clases. La metodología de amplio margen consigue intuitivamente la mayor robustez frente a perturbaciones tales como el ruido y los outliers. Por otro lado, en SVM, el riesgo de los outliers se mitiga normalmente utilizando un criterio de margen suave, de la misma forma que la función de pérdidas de LMNN se utiliza para reducir la sensibilidad de los outliers.

5.6 Métodos para la evaluación

Las técnicas de validación están motivadas por dos principios fundamentales: la selección del modelo y la estimación del rendimiento. Los métodos de evaluación utilizados en este trabajo se clasifican en: validación cruzada, dejar-uno-fuera y K -pliegues. La validación cruzada es una metodología estadística para evaluar y comparar algoritmos de aprendizaje dividiendo los datos en dos segmentos: uno utilizado para el aprendizaje o entrenamiento de un modelo (grupo de entrenamiento) y el otro para validar el modelo (grupo de test). En la validación cruzada típica, los conjuntos de entrenamiento y validación deben intercambiarse en sucesivas vueltas de tal forma que cada elemento de la base de datos tenga la oportunidad de ser validado [Refaeilzadeh et al., 2008], [I.A.Illán, 2009]. Otro objetivo de la validación cruzada consiste en evitar el sobreajuste dado que los elementos pertenecientes al conjunto de test no pueden pertenecer al conjunto de entrenamiento para que haya un ajuste adecuado del modelo a los datos de prueba igual de bueno que a los datos de entrenamiento.

En validación cruzada K -pliegues, a veces llamada estimación rotativa, el conjunto de datos D se separa aleatoriamente en K subconjuntos exclusivos (los pliegues) D_1, D_2, \dots, D_k de aproximadamente el mismo tamaño, como se muestra en la figura 5.6. Cada elemento de la base de datos se entrena y

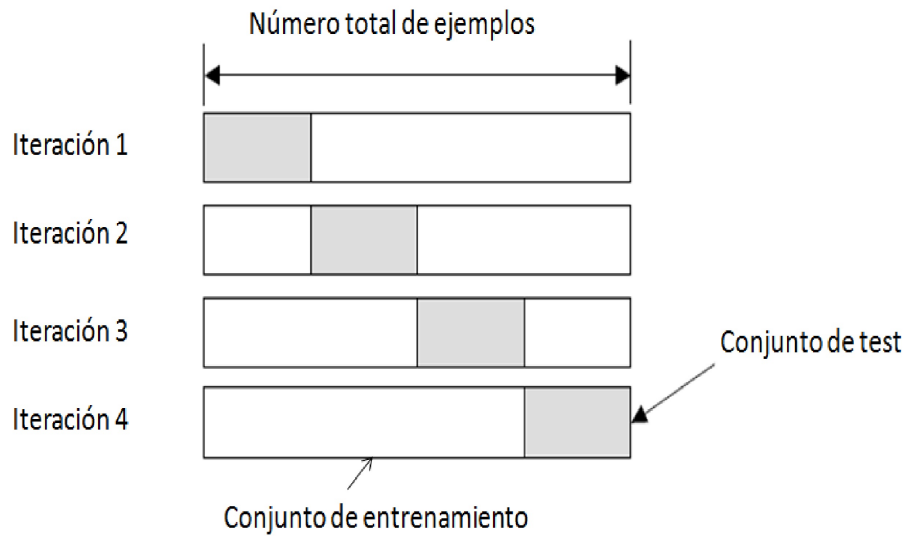


Figura 5.6: Validación cruzada de k pliegues, para K=4

testea K veces; cada vez $t \in \{1, 2, \dots, K\}$, se entrena en D/D_t y se testea en D_t . La estimación de la exactitud de la validación cruzada es el número de clasificaciones correctas en general, divididas por el número de instancias de la base de datos. La estimación de la validación cruzada es un número aleatorio que depende de la división en pliegues, siendo $k = 10$ una elección muy común. La validación cruzada *completa* es el promedio de todas las

$$\binom{m}{m/k}$$

posibilidades para elegir m/k instancias fuera de m , pero requiere un elevado coste computacional. La validación cruzada de k pliegues completa estima utilizando una única separación de los datos en pliegues. Repetir la validación cruzada múltiples veces usando diferentes separaciones en pliegues proporciona una mejor estimación de Monte-Carlo para la validación cruzada completa pero a un coste adicional. La validación cruzada dejar uno fuera (LOO) es un caso especial de validación cruzada completa donde k iguala al número de instancias de los datos. En otras palabras, en cada iteración se entrena con todos los datos excepto con el que se evalúa. LOO se considera una validación cruzada de n pliegues (número total de instancias).

Las medidas utilizadas en este trabajo para medir la bondad del método de clasificación son la exactitud (Acc , del inglés “Accuracy”), sensibilidad (Sen , del inglés “Sensibility”) y especificidad (Spe , del inglés “Specificity”). La Acc es la probabilidad de clasificar correctamente una instancia seleccionada aleatoriamente y se define como:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (5.43)$$

La Sen y Spe de cada sujeto de test se definen como:

$$Sen = \frac{TP}{TP + FN}; \quad Spe = \frac{TN}{TN + FP}$$

respectivamente, donde TP es el número de verdaderos positivos (número de pacientes AD correctamente clasificados); FP es el número de falsos positivos (número de controles clasificados como pacientes AD); FN es el número de falsos negativos: número de pacientes AD clasificados como controles y TN es el número de verdaderos negativos: número de controles clasificados como controles. Otra forma usada en esta tesis para mostrar la bondad de los métodos de clasificación es mediante los análisis de la curva operativa del receptor (ROC, del inglés “Receiver Operating Characteristic”) y el área bajo la curva (AUC, del inglés “Area Under Curve”).

5.6.1 Análisis ROC y AUC

La curva ROC fue diseñada para las pruebas de diagnóstico, dado que describe las características de detección inherentes del test. En ella se representa el compromiso entre los pares de coordenadas $(1 - spe, sen)$ de los verdaderos positivos frente a los falsos positivos conforme varía un umbral de decisión y es independiente de la prevalencia de la enfermedad en la población a la que el test se aplica [Metz, 1978]. Los ejes de la ROC pasan de 0 a 1, dado que son los límites que $(1 - spe, sen)$ pueden adquirir. La figura 5.7 muestra los tres puntos posibles de operación que pueden corresponderse al umbral estricto (es el punto de operación de la curva, es decir el más arriba y a la izquierda dado que la especificidad y sensibilidad son máximas a la vez y se pronostica caso positivo de enfermedad sólo si se juzgó definitivamente positivo), al umbral moderado o finalmente al umbral relajado (caso positivo de enfermedad si hay cualquier sospecha de la misma, dado que la sensibilidad es elevada).

Otra forma de ver la bondad de un método de clasificación, es calculando la AUC [Han and Zhao, 2010]. Un área de 1, representa un test perfecto, un área de 0.5 representa un test sin valor.

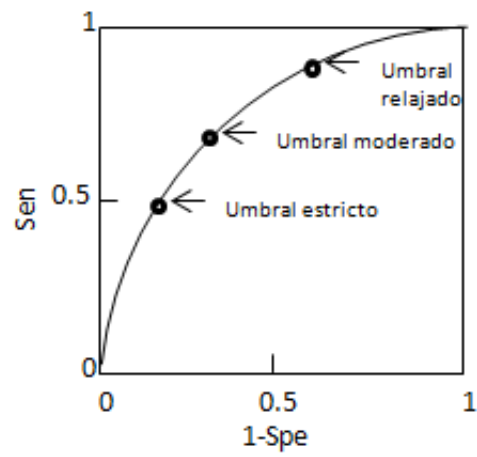


Figura 5.7: Una curva ROC convencional, se muestran los tres puntos de operación posibles. Fuente [Metz, 1978].

Parte IV

Análisis de imagen con los métodos propuestos

Capítulo 6

Análisis de imagen mediante ARs

Aunque las ARs fueron en su origen utilizadas en el análisis de mercados, su aplicación a imagen médica resulta tan novedosa como eficaz. Este capítulo comienza introduciendo diferentes ejemplos en los que se ha combinado el uso de las ARs con imagen médica. Siguiendo esta línea, proponemos la aplicación de las ARs como metodología válida y novedosa para el análisis de imagen funcional PET y SPECT para el diagnóstico precoz de la AD desde diversas perspectivas: i) entre distintos tipos de atributos (continuos, discriminantes y/o discretos) e ii) como selector de características seguido de la extracción de las mismas proporcionada por PCA o PLS. Los métodos propuestos se describen en sus fases de enmascarado, obtención de ROIs, extracción de ARs (entre las ROIs del grupo control) y clasificación según el número de reglas verificadas por cada sujeto (en el primer caso i)) o mediante un kernel SVM (en el segundo ii)) siguiendo en ambos una evaluación LOO.

6.1 Análisis de imagen médica con ARs

Aunque las ARs han sido típicamente usadas para el análisis de mercados, han sido también propuestas para el análisis y categorización de imágenes con resultados prometedores. En [Ordoñez and Omiecinski, 1998] se analiza el problema de encontrar asociaciones en imágenes a color bidimensionales en 4 pasos: extracción de características, identificación del objeto, creación de la imagen auxiliar y extracción del objeto para producir ARs. El algoritmo básico para encontrar asociaciones entre imágenes se basa en la búsqueda de similitud entre imágenes sintéticas con descriptores basados en una función de distancia.

Las ARs han sido propuestas también como una alternativa para la clasificación de imágenes médicas. En [Ordoñez et al., 2000], la base de datos con la que se trabaja describe los perfiles de enfermos de corazón tratados en un hospital. Aparte de información personal del paciente como la edad, si es o no fumador, se incluyen medidas como el peso, presión arterial, o factores como si padece o ha padecido otra enfermedad. También se almacena la información de imagen de varias regiones del corazón. Además se concluye con que muchas ARs obtenidas pueden no ser útiles porque contienen información redundante, son triviales, muy complejas o simplemente no tienen un significado médico. Por ejemplo, si el médico desea relacionar la edad con los hábitos de fumar entonces la regla puede ser útil, pero si sólo está intentando conocer el impacto de estos dos factores conjuntamente en la enfermedad del corazón, entonces la regla es irrelevante.

Las ARs han supuesto también un avance en el estudio del cáncer. En [Antonie et al., 2001] se estudia la clasificación de imágenes médicas de mamografías para la detección de cáncer de mama, definiendo las ARs de tal forma que el antecedente consiste en una conjunción de características de la mamografía mientras que el consecuente de la regla es la categoría a la que la mamografía pertenece. Cuando una nueva imagen tiene que ser clasificada, el sistema de categorización devuelve las ARs que aplica a esa imagen. Las ARs estadísticas son empleadas en [Ribeiro et al., 2006] para seleccionar las características más relevantes de bases de datos de mamografía y heterogéneas. De este modo se pueden discriminar imágenes médicas, reduciendo el tamaño de los vectores de características y eliminando las características ruidosas que influyen negativamente en los resultados, haciendo el proceso completo más eficiente. En [Zaiane et al., 2002] se introduce un sistema de clasificación para la detección temprana del cáncer de mama basado en la extracción de reglas por categorías. Extrae el conjunto de datos de imágenes de mamografías organizadas por clases en lugar de extraerlos todos conjuntamente. Dada una

nueva imagen, el proceso de clasificación busca en este conjunto de reglas para encontrar la clase que mayor número de reglas verifica. En [Zubi and Saad, 2011] se lleva a cabo una detección temprana del cáncer de pulmón mediante la clasificación con reglas de asociación y redes neuronales aplicadas a radiografías del tórax. Las ARs se pueden expresar de dos formas:

1. Los contenidos de una imagen son diferentes a las relaciones espaciales: por ejemplo si una imagen tiene textura X , es probable que contenga protusión Y .
2. Los contenidos de una imagen están asociados con las relaciones espaciales: por ejemplo si X está entre Y y Z , es probable que exista un T debajo.

Finalmente, las ARs permiten construir clasificadores asociativos como los que se describen en la sección 5.5.1. En [Liu et al., 1998a] tras discretizar atributos continuos se construye un clasificador asociativo, es decir, se integra clasificación y extracción de reglas. De esta forma no sólo se extraen las reglas de un conjunto de items, sino que también se tiene en cuenta la clase a la que pertenecen. En [Sheela and Shanthi, 2009] se propone un algoritmo supervisado de preprocesado para ARs que lleva a cabo la discretización de valores continuos de características. Del mismo modo integra extracción de reglas de asociación y de clasificación ya que se centra en un tipo especial de ARs llamadas reglas de clasificación que asocian un ítem a una clase con un soporte y confianza. A continuación se presentan los métodos propuestos en este trabajo, clasificándolos en las secciones: ARs entre atributos continuos, discriminativos o discretos y ARs como técnica de selección de características.

6.2 ARs entre atributos continuos

En el mundo real muchos atributos son por naturaleza continuos, por ejemplo: la altura, el peso, la longitud, la temperatura y/o la velocidad. Manejar adecuadamente este tipo de datos es esencial para trabajar con un sistema de minería de datos. En algunos casos los algoritmos pueden ser adaptados para ser utilizados con atributos continuos. En otros casos esto no es posible. Aunque sería posible tratar un atributo continuo como uno categórico con valores (por ejemplo 6.3, 7.2, etc), puede tener resultados poco satisfactorios en general. Si el atributo continuo tiene muchos valores diferentes en el conjunto de entrenamiento, es probable que cualquier valor particular ocurra pocas veces (quizá sólo una), y las reglas que incluyen los tests para valores específicos tales como $X = 7.2$ tengan poco valor de predicción [Fayyad and Irani, 1993].

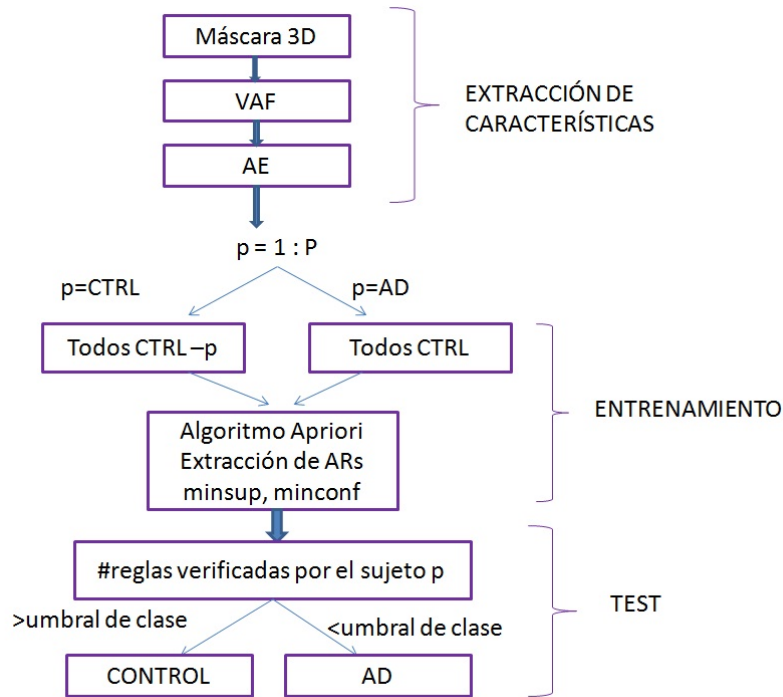


Figura 6.1: Esquema con la extracción de características, entrenamiento y test del sistema basado en ARs para atributos continuos.

La importancia de las relaciones entre áreas cerebrales encargadas de la memoria a largo plazo y de enviar información a las cortezas prefrontales ejecutivas tales como la corteza prefrontal medial, la orbitofrontal y la cingulada anterior [Contreras et al., 2008] se pone de manifiesto en este capítulo para la obtención de atributos continuos, dado que no podrían encontrarse a nivel de vóxel o en un análisis ROI directo. Así el primer método que proponemos representa cada paciente por un vector de características que contiene las ROIs equivalentes a bloques tridimensionales con una alta concentración de coordenadas de activación [Newman et al., 2008], [Chaves et al., 2010] relacionadas directamente con el nivel de perfusión que varía ampliamente entre controles y AD en sus 3 posibles estadios. Finalmente, se establecen relaciones entre ROIs en términos de reglas de asociación para distinguir eficientemente entre pacientes AD y controles. La descripción del método basado en ARs para atributos continuos se detalla en la figura 6.1 y se desglosa en tres fases: i) extracción de características, ii) entrenamiento, y iii) test [Chaves et al., 2011a].

6.2.1 Extracción y selección de características

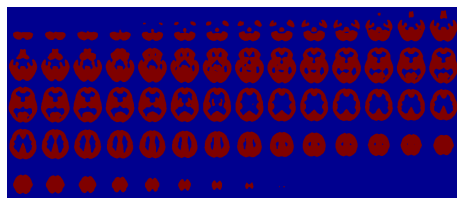
Se propone aplicar una combinación de VAF y de la estimación de la activación basada en umbral (AE, del inglés activation estimation), un método similar al descrito por [Turkeltaub et al., 2002] para la extracción de características de cada sujeto. El método base (VAF) [Stoeckel et al., 2004] es una forma de incluir todas las intensidades dentro del cerebro como características, de forma que no se necesita un conocimiento previo de la enfermedad y se evita por tanto la inclusión de información *a priori* en el sistema. Esto equivale a que la información se obtiene implícitamente de las bases de datos de las imágenes.

Enmascarado en atributos continuos

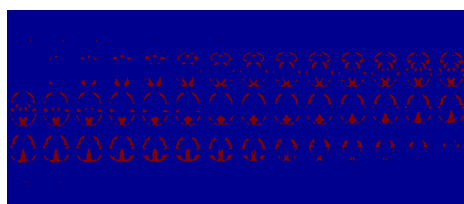
En imagen funcional, cada voxel en el cerebro de coordenadas (x, y, z) de un paciente posee un nivel de intensidad $I(x, y, z)$, y dependiendo de la modalidad de adquisición de la imagen, está relacionado con el flujo sanguíneo cerebral [Holman et al., 1992] en SPECT, o con el metabolismo de la glucosa en PET [Ng et al., 2007]. En el primer paso, se calcula una imagen tridimensional media a partir de todas las imágenes de controles de la base de datos. Se fija un umbral de activación a_T del 50% de la intensidad máxima de esta imagen media, definiendo de este modo una máscara 3D. Un voxel se considera como *activado* cuando su valor de intensidad I está por encima de un umbral de activación, es decir, cuando $I > a_T$. La figura 6.2 muestra los cortes 3D de las máscaras tridimensionales para las bases de datos SPECT, FDG-PET (ADNI) y PET-Cartuja que se describen en el capítulo 8. Siguiendo la metodología de VAF [Stoeckel et al., 2004], los vóxeles que se encuentren fuera de esta máscara se descartan, excluyéndose del análisis las regiones que quedan fuera del cerebro y las regiones poco activadas. De esta forma se define una matriz VAF (**vaf**) de dimensión número de pacientes (P) por número de vóxeles considerados en las coordenadas de la máscara ($mask(x, y, z)$) como:

$$\mathbf{vaf} = (\mathbf{I}_1(mask(x, y, z)), \dots, \mathbf{I}_P(mask(x, y, z))) \quad (6.1)$$

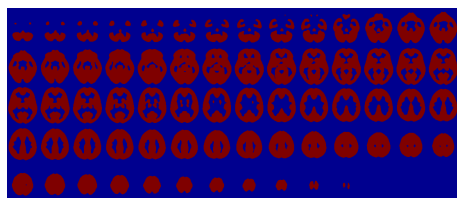
La figura 6.3 muestra un corte coronal de la base de datos SPECT con las fases de preprocesado descritas previamente: a) SPECT original, b) promedio de controles, c) máscara con $I > a_T$, d) VAF en las coordenadas de la máscara. El principal problema de utilizar VAF es *la maldición de la dimensionalidad*, dado que el número de muestras disponibles es mucho menor que el número de características usadas en el entrenamiento. Es por ello que en este capítulo, para solventar este problema utilizamos la estimación de la activación descrita a continuación.



(a)



(b)



(c)

Figura 6.2: Cortes coronales de la máscara 3D (basada en el promedio de controles) para: a)SPECT, b)FDG-PET(ADNI), c)PET-Cartuja

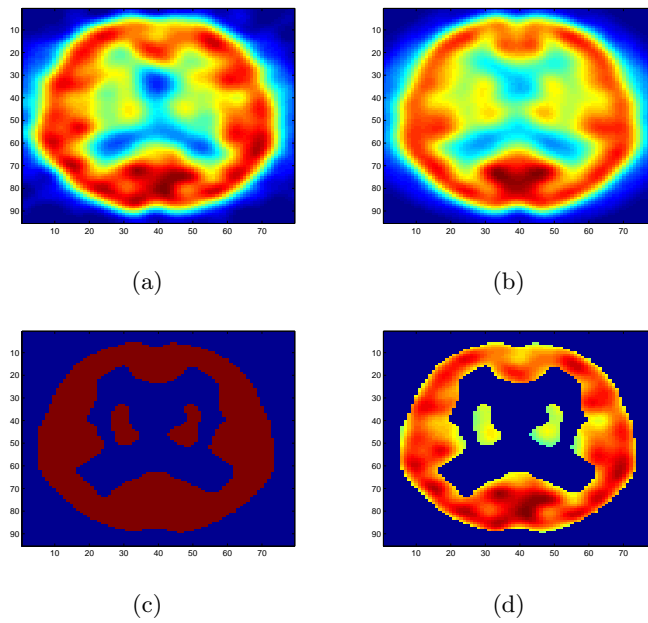


Figura 6.3: Corte coronal SPECT para mostrar el emascarado basado en el promedio de normales en las fases de: a) SPECT original, b) promedio de controles, c) máscara, d) VAF: corte emascarado

Estimación de la activación

Respecto a la AE, en [Turkeltaub et al., 2002] se introdujo la utilidad de identificar áreas de activación consistentes mediante métodos meta-analíticos (integran o reutilizan la información ya obtenida para poder obtener un análisis mejor) para identificar procesos asociados con la cognición y sensorimotora. En este trabajo, la AE calcula una lista reducida de ROIs con activación máxima que contienen un conjunto de vóxeles activados. En la fase de AE, cada imagen se divide en vóxeles cúbicos tridimensionales $v \times v \times v$ como en [Chaves et al., 2009b], o bloques centrados en las coordenadas válidas obtenidas con el enmascarado. El número de bloques en los que una imagen se divide depende de dos parámetros: v y g (véase figura 6.4). El primer parámetro define el tamaño de los bloques, y el segundo define el solapamiento o número de vóxeles entre dos bloques adyacentes en una dirección. Las localizaciones del punto central de los bloques definen una malla tridimensional de tamaño $g \times g \times g$, considerando la posibilidad de solapamiento entre bloques adyacentes. Un bloque se parametriza por la proporción de vóxeles activados t . Se ha considerado que un bloque está activado sólo si el ratio de vóxeles activados dentro de él es superior a un umbral dado t , es decir:

$$\left(\sum_{x_m, y_m, z_m} I(x_m - v' : x_m + v', y_m - v' : y_m + v', z_m - v' : z_m + v') > a_T \right) > t \quad (6.2)$$

donde x_m, y_m, z_m definen las coordenadas válidas de la máscara y v' se define como la mitad del tamaño del vóxel. Estos parámetros permiten reducir el coste computacional [Tan et al., 2005]. Si el umbral de activación se reduce o se incrementa el solapamiento, el coste computacional del sistema aumenta.

Por otro lado, se analizaron otras estrategias para reducir el coste computacional de nuestra aproximación, añadiendo restricciones extra a la estructura de los patrones [Umarani and Punithavalli, 2010]. En este sentido, usamos una decisión basada en umbral que provee un buen compromiso entre coste computacional y exactitud y permite la inclusión de todas las regiones cerebrales relevantes que se ven afectadas en la progresión de la AD.

6.2.2 Extracción de las reglas de asociación

Como se describió en el capítulo 5 de este trabajo, el aprendizaje de reglas de asociación es un método efectivo para descubrir relaciones interesantes entre variables en bases de datos grandes [Agrawal et al., 1993]. En este contexto, el número de items m representa el número de regiones activadas (también llamadas ROIs) $I = i_1, i_2, \dots, i_m$ y T define una base de datos de transacciones

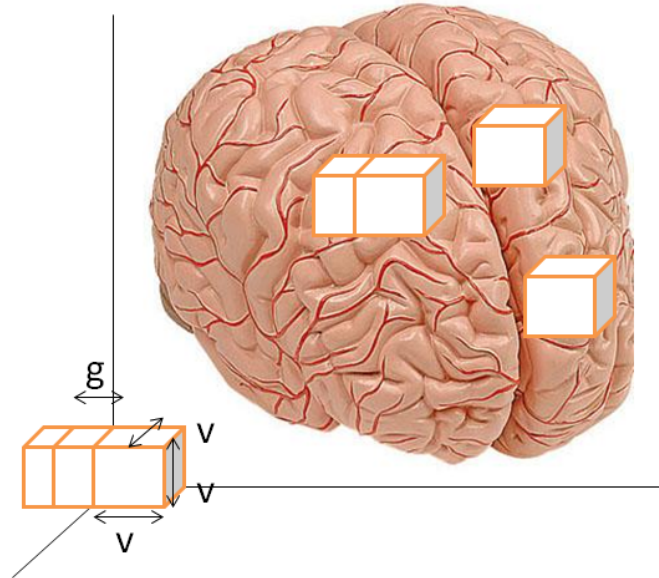


Figura 6.4: Definición de bloque tridimensional, con parámetro v para definir su tamaño y g para su solapamiento.

$T = t_1, t_2, \dots, t_n$ que relacionan sólo las ROIs de los controles. Cada transacción t está representada como $X \Rightarrow Y$, donde $X, Y \subset \text{ROIs de controles}$ y $X \cap Y = \emptyset$. En nuestro problema, una regla de asociación $X \Rightarrow Y$, que relaciona dos items activados (regiones) del patrón normal, se define como la activación en el área X que implica activación en el área Y , dadas ciertas restricciones de soporte y confianza. La tabla 6.1 representa la entrada al algoritmo *Apriori*. Las ROIs de cada control se representan binariamente (como se describió en el capítulo 5) como activado o no activado, según si supera o no el umbral de activación t respectivamente.

La extracción de ARs se lleva a cabo entre los patrones normales dado que estos son menos variables que los patrones de AD [Illán et al., 2010], por lo que son más adecuados para caracterizar las relaciones “AR” relevantes entre regiones cerebrales, que no se satisfarán en general por los pacientes AD. Además, la extracción de las ARs sólo de los controles resuelve el problema de la alta dimensionalidad descrito en [Illán et al., 2011a]. Esta metodología mejora los resultados de clasificación ya que tenemos diferentes muestras del mismo individuo, es decir, el sistema CAD propuesto basado en AR sigue la filosofía de las aproximaciones estadísticas relacionadas con el voxel, como por

Controles	ROI 1	ROI 2	ROI 3	ROI 4	...	ROI N
1	1	1	0	0	...	0
2	1	0	1	1	...	0
3	0	1	1	1	...	1
4	1	1	1	1	...	0
...	1
N	1	1	1	0	...	1

Tabla 6.1: Ejemplo de representación binaria de imágenes funcionales. Fuente [Tan et al., 2005]

ejemplo en SPM [Friston et al., 2007]. Es decir, en la práctica, cada individuo de la base de datos se representa por una serie de ROIs y, como consecuencia, se mitiga el problema del pequeño tamaño muestral [Duin, 2000].

El problema de la extracción de ARs consiste en generar todas las reglas que tienen soporte y confianza superior a un mínimo-soporte (*minsup*) y a una mínima-confianza (*minconf*), respectivamente, establecidas por el usuario. El algoritmo *Apriori* [Agrawal and Srikant, 1994] se usa para la extracción de ARs. Si los umbrales de *minsup* y *minconf* se incrementan, el número de reglas extraídas se reduce y vice-versa. En particular, si *minconf* y *minsup* se seleccionan cercanas al máximo del 100%, el número de reglas extraídas es menor y sólo las características más discriminativas se usan en la etapa de clasificación. De esta manera, se extraen las reglas que mejor distinguen los sujetos AD de los controles. Generalmente, las regiones cerebrales temporoparietales, que son regiones características cerebrales que se ven alteradas en la progresión de la AD, se consideran como activadas por el algoritmo [Fung and Stoeckel, 2007].

6.2.3 Clasificación

En la fase de test, se clasifica cada paciente en función del número de reglas que verifica como se observa en la figura 6.1 y se explica con detalle en la sección 5.5.1. Dado que las ARs se extraen de los controles, se espera que los sujetos sanos verifiquen un número más elevado de ARs que los AD.

- Fase de entrenamiento: Las ROIs fueron previamente obtenidas para cada sujeto de la base de datos como se describe en la sección 6.2.1. Las ARs se extraen de los controles en un proceso iterativo LOO (método de evaluación detallado en la sección 5.6) es decir, el control que se evalúa no se tiene en cuenta en la extracción de reglas.

- Fase de test: Se comprueba el número de ARs que son verificadas por cada sujeto de test siguiendo la estrategia de validación LOO. Si un sujeto verifica un número mayor o igual de ARs que un *umbral de clase* entonces se clasifica como normal, en otro caso como AD.

6.3 ARs entre atributos discriminativos

En esta sección, se asocian las ARs entre regiones discriminativas con dos motivaciones fundamentales [Chaves et al., 2012c]: reducir el coste computacional del sistema basado en reglas y generalizar su funcionamiento en distintas bases de datos (SPECT y FDG-PET de ADNI) utilizando para ello los mismos parámetros: tamaño de bloque, solapamiento entre regiones y umbral FDR. Cada vóxel de una imagen tridimensional SPECT y PET contiene información acerca del flujo sanguíneo (rCBF) o del metabolismo de glucosa del área cerebral correspondiente respectivamente. Sin embargo, no todos los vóxeles tienen el mismo nivel de relevancia en términos de discriminación entre grupos de sujetos (CTRL frente a AD). Por lo tanto, una selección de vóxel inicial basada en la capacidad de discernir entre clases con el cálculo de FDR (descrito teóricamente en la sección 5.1.1) [Illán et al., 2011b; Ramírez et al., 2009b] permite obtener un vector de vóxeles discriminativos para cada participante con el objetivo de mejorar el coste computacional de las ARs entre atributos continuos.

6.3.1 Descripción del método

La figura 6.5 muestra un diagrama de bloques del método propuesto que consiste en dos fases:

1. Fase de entrenamiento para extraer las ARs: introduce la selección de vóxeles discriminativos obtenidos al pasar una máscara basada en FDR (como se explica en la sección 6.3.2). El proceso de entrenamiento lleva a cabo la estimación de la activación (para la selección de vóxeles) y la extracción de reglas utilizando el conjunto de controles de la misma forma que se describió en las secciones 6.2.1 y 6.2.2.
2. Fase de test para la clasificación de un sujeto bajo estudio (que no se considera en la fase de entrenamiento) basado en las reglas previas extraídas como se describió en la sección 6.2.3.

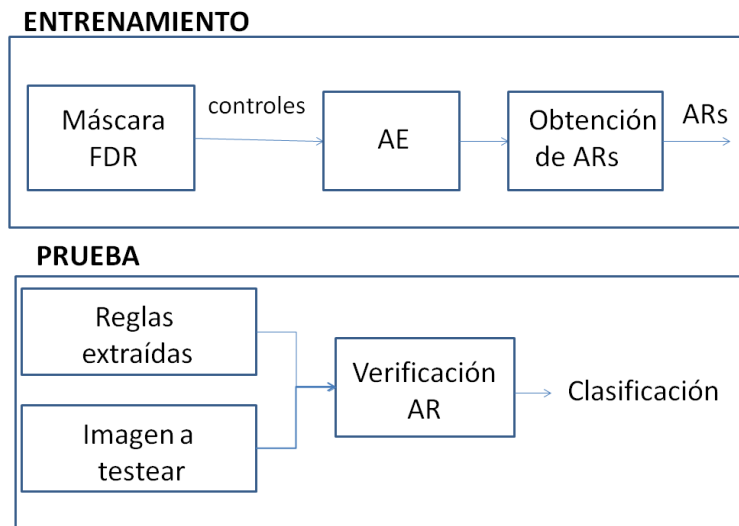


Figura 6.5: Diagrama de método de ARs sobre regiones discriminativas

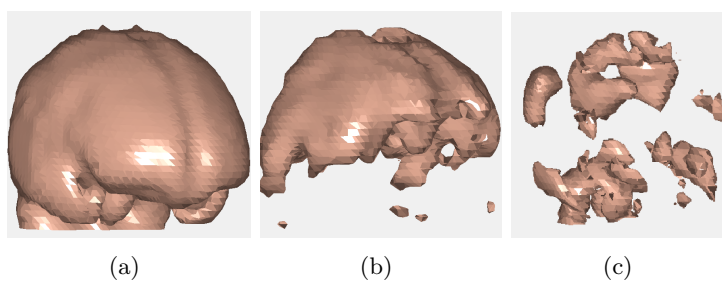


Figura 6.6: Máscara 3D: a) promedio de controles para la base de datos SPECT, b) máscara FDR SPECT, c) máscara FDR para PET.

6.3.2 Enmascarado

El proceso de enmascarado FDR se realiza para identificar los vóxeles discriminativos para la extracción de ARs. Este paso no sólo mejora la exactitud del método propuesto sino que también reduce el gasto computacional [Ramírez et al., 2009b]. Además, los vóxeles discriminativos pueden usarse como centros de los bloques cúbicos con un solapamiento entre los situados adyacentes para su posterior procesado. La figura 6.6 muestra las máscaras para a) el promedio de controles como referencia, b) la máscara SPECT basada en FDR y c) la máscara FDR basada en PET. Las últimas imágenes se obtuvieron a un umbral de FDR denominado h igual a 0.2 (como se explica en el capítulo 9 de experimentos), es decir, sólo los vóxeles con un umbral FDR por encima de 0.2 del máximo se tienen en cuenta para un análisis posterior. Se destaca que la figura 6.6(a) muestra todos los vóxeles que están dentro del cráneo para que se consideren en la extracción de ARs mientras que en las figuras 6.6(b) y 6.6(c), se muestran las regiones específicas, descritas por el criterio FDR y usadas para extraer las ARs después. Esta reducción de la dimensión provee buenos resultados de clasificación junto con un gasto computacional menor a la vista de los resultados experimentales (véase capítulo 9 de experimentación). La clasificación se llevó a cabo por el método basado en reglas descrito en la sección 6.2.3.

6.4 ARs entre atributos discretos

Convertir un atributo continuo en uno discreto, por ejemplo un atributo categórico, se denomina discretización (véase sección 5.2.4). Los datos discretizados apropiadamente pueden simplificar el proceso de aprendizaje y mejorar la generalidad de los resultados aprendidos como se describió en el capítulo 5 de teoría.

La discretización se ha utilizado en el campo de la neuroimagen, por ejemplo en [Ostwald and Bagshaw, 2011] donde los conceptos de la teoría de la información tales como la entropía [Cabella et al., 2009] (para medir la incertidumbre de una fuente de información) se aplican a imagen funcional. La estimación de la entropía se lleva a cabo en este caso mediante la discretización de los valores de los vóxeles con un tamaño constante de intervalo. Encontramos otro ejemplo en [Chen and Herskovits, 2010] donde se investiga el diseño de modelos de diagnóstico para diferenciar a los individuos con demencia temprana de la gente anciana utilizando imágenes MRI. En particular se incide en el clasificador de redes bayesianas Naïve Bayes basado en variables categóricas, que utiliza el método descrito en [Fayyad and Irani, 1993] para

discretizar cada variable continua en binaria. En general, este método puede usarse en cualquier algoritmo para el aprendizaje de árboles de clasificación o reglas que manejen atributos de valor continuo mediante la cuantización de sus rangos en dos intervalos.

En las secciones anteriores se han considerado atributos continuos y discriminantes para la extracción de ARs. En este capítulo las reglas se extraen de las ROIs de valor discretizado [Chaves et al., 2013]. En primer lugar la elección de la máscara se lleva a cabo mediante un histograma de imagen y en segundo lugar las ROIs que se introducen como entrada al algoritmo *A priori* se seleccionan por una combinación de “bins” discretizados a partir de la intensidad media. Hay numerosas aplicaciones para discretizar atributos continuos. Idealmente los puntos límite elegidos para los rangos (los puntos de corte) deben reflejar las propiedades reales del dominio que se investiga. En la práctica es difícil dar las razones principales para elegir un conjunto de rangos frente a otro. Los experimentos (véase capítulo 9) se generalizaron para las bases de datos SPECT y FDG-PET en lo que a histograma de imagen (para el enmascarado) o número de intervalos de discretización se refiere, obteniendo resultados mejorados respecto al caso continuo.

6.4.1 Descripción del método

El algoritmo propuesto combina en un único paso la selección de características y la discretización como en [Ribeiro et al., 2009]. Esta adaptación conlleva discretizar atributos continuos (la intensidad media de vóxel 3D en nuestro caso) en intervalos de igual tamaño de intervalo (véase sección 5.2.4 para detalles teóricos). Si el nivel medio de activación de una característica pertenece al intervalo (o combinación de intervalos) de un intervalo (o intervalos respectivamente) discretizados, esta característica se considera como ROI para la extracción de ARs.

Este capítulo muestra un método novedoso para encontrar asociaciones entre atributos continuos de imágenes funcionales mediante: i) examinar las diferencias y analogías entre bases de datos, ii) seleccionar los mejores intervalos con el fin de eliminar características irrelevantes para el diagnóstico precoz de la AD. Se describe el sistema CAD que combina discretización y ARs en detalle [Chaves et al., 2013]. En primer lugar, se lleva a cabo el preprocesamiento de SPECT y PET, mediante el histograma de imagen para el procedimiento de enmascarado (sección 6.4.2) seguido de la discretización (para obtener las ROIs entrada del algoritmo de extracción de ARs) descrita en la sección 6.4.3, hasta la clasificación basada en ARs de los sujetos en dos categorías (clases CTRL(0) y AD(1)) mediante el método descrito en la sección 6.2.3. La figura

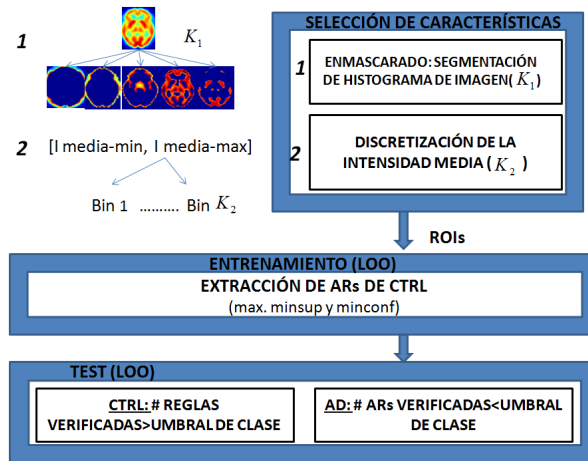


Figura 6.7: Esquema con las fases del sistema CAD propuesto: selección de características, entrenamiento y test.

6.7 muestra un diagrama de bloques del sistema propuesto que consiste en 4 etapas: *i*) procedimiento de enmascarado (sección 6.4.2), *ii*) discretización de la intensidad media para la selección de características (véase sección 6.4.3), *iii*) extracción de ARs (como se describió en la sección 6.2.2), *iv*) clasificación LOO con el método basado en reglas (el mismo que el descrito en sección 6.2.3 para los atributos continuos y discriminantes).

6.4.2 Enmascarado

En primer lugar, se calcula una imagen media de controles y un histograma de imagen se aplica sobre ella para segmentarla en K_1 subimágenes que contienen los vóxeles cuyas intensidades pertenecen a cada intervalo. A esto también se le denomina umbralizar la imagen y está basada en la suposición de que los objetos pueden distinguirse por su nivel de intensidad [Nakib et al., 2007]. La máscara descarta los vóxeles de la subimagen con un valor de cero (representado en azul en la figura 6.8). Del total de K_1 intervalos, si se seleccionan los mayores para la obtención de la máscara, se obtienen mejores resultados para la detección precoz de AD, mientras que si se seleccionan los menores se introduce ruido en el sistema CAD. La figura 6.8 muestra (en la primera fila para SPECT y en la segunda para PET) los cortes transaxiales correspondientes a las máscaras obtenidas del promedio de los normales (representadas en figuras

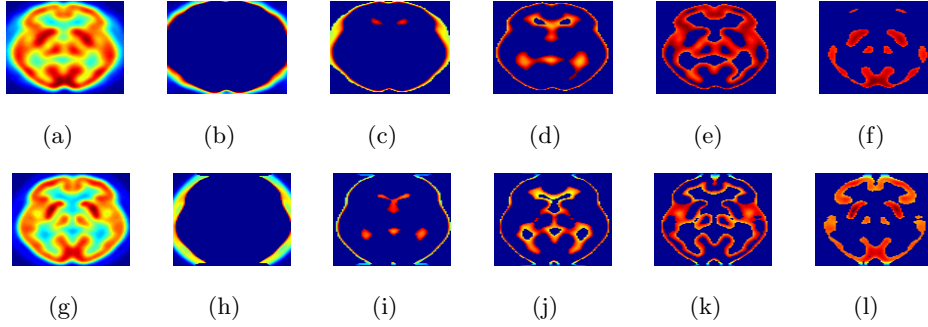


Figura 6.8: Cortes transaxiales de la segmentación del histograma de imagen (SPECT en la primera fila y PET en la segunda) obtenidos de las imágenes promedio control (a) y g)) que se dividen respectivamente en las siguientes $K_1 = 5$ subimágenes correspondientes a cada intervalo.

6.8(a) y 6.8(g) respectivamente) para ambas bases de datos cuando se segmentan las $K_1 = 5$ subimágenes en términos de un histograma de intensidad (véanse figuras 6.8(b-f) y 6.8(h-l)). Se observa que las máscaras seleccionadas (la cuarta para SPECT mostrada en la figura 6.8(e) y la quinta para PET en 6.8(l)) incluyen los vóxeles más relevantes para el diagnóstico de la AD ya que se sitúan en las regiones temporoparietales. La razón por la que el cuarto intervalo se seleccionó para SPECT y el quinto para PET se explica en la sección de *Experimentos con base de datos PET* del capítulo 9 y se relaciona con el nivel de activación.

6.4.3 Vóxeles seleccionados por discretización

Las coordenadas de la máscara son el centro de los bloques cúbicos tridimensionales. Cada bloque se representa por una característica continua que consiste en la intensidad media (I_{mean}) de los vóxeles cuya intensidad pertenece al rango de la máscara $[I_{min}, I_{max}]$. La discretización consiste en un histograma de igual anchura aplicado al rango $[I_{mean-min}, I_{mean-max}]$ dividiéndolo en K_2 intervalos (en inglés “bins”) de igual anchura (véase más información en la sección 5.2.4 y en [Dougherty et al., 1995]). Los intervalos discretizados se usan independientemente o de forma combinada para definir los intervalos en los que el valor de la ROI seleccionada está para cada sujeto. Estas ROIs actúan como entradas para el algoritmo *Apriori* de extracción de ARs. Se demuestra que las combinaciones de intervalos discretizados más bajos suelen ser ruidosas para clasificación, mientras que las más altas tienen alta proba-

bilidad de mostrar información relevante para la detección de AD temprano (véase sección 9.3). La extracción se lleva a cabo con *Apriori* utilizando *minsup* y *minconf* de 100% a partir de los controles y siguiendo la metodología LOO por el mismo proceso que el descrito en el para ARs sobre atributos continuos y discriminativos, según el umbral de reglas verificadas por cada paciente en la sección 6.2. Una AR discretizada en [Ribeiro et al., 2009] se define como $R_1[a, b]R_2[c, d] \rightarrow clase_1$ que quiere decir que dadas las características R_1, R_2, \dots, R_N extraídas para cada paciente, podemos decir que si para un paciente la característica R_1 tiene un valor comprendido en el intervalo $[a, b]$ y la característica R_2 en el intervalo $[c, d]$ implica su pertenencia a la clase 1. Esta forma de definir las ARs discretizadas con combinación de intervalos está relacionada con la utilizada en esta Tesis. La diferencia radica en que las reglas de esta tesis las extraemos únicamente de los controles ya que tienen un patrón más continuo para detectar AD. Un regla podría definirse como: $R_1[a, b] \rightarrow R_2[c, d]$, que quiere decir que para verificar una regla discretizada que define el patrón normal, el sujeto debería verificar que su valor de característica R_1 perteneciese al intervalo $[a, b]$ y que el valor de característica R_2 perteneciese al intervalo $[c, d]$. La clasificación se logra por el método basado en reglas descrito en la sección 6.2.3.

6.5 ARs como selector de características

En las secciones anteriores se han analizado las ARs sobre atributos continuos, discriminantes y discretos para mejorar el coste computacional presentándose un clasificador basado en reglas. En esta sección las reglas de asociación se utilizan como selector de características [Chaves et al., 2012e] condicionados a parámetros como la activación de las ROIs o los umbrales *minsup* y *minconf* del algoritmo *Apriori* para la extracción de ARs. De esa elección dependerá la reducción del espacio de características del cual se extraen las más adecuadas mediante técnicas de extracción PCA o PLS. Por último la clasificación se lleva a cabo mediante kernel SVM. El método se aplica en la base de datos SPECT y FDG-PET obteniendo en ambos casos resultados de clasificación por encima del 90%.

6.5.1 Descripción del método

A continuación se propone un método de FS basado en ARs, similar al descrito en [Karabatak and Ince, 2009], combinado con técnicas de FE para reducir el espacio de entrada a un clasificador SVM para el diagnóstico precoz de la AD. La descripción del método se representa en la figura 6.9 que consta

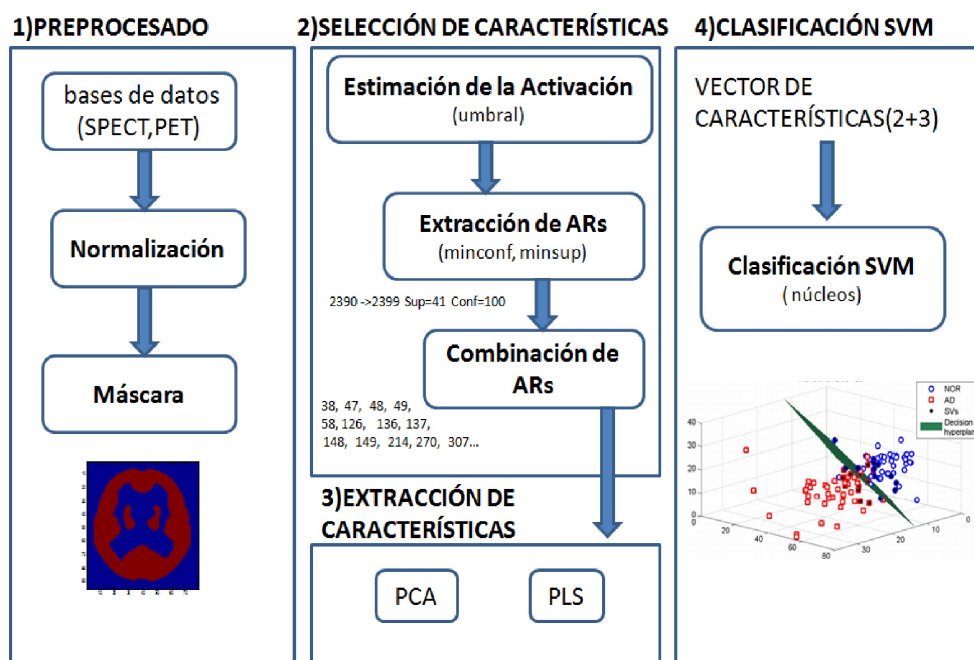


Figura 6.9: Diagrama explicativo del sistema CAD de selección de características basado en ARs con el preprocesado, selección y extracción de características y clasificación.

de preprocesado, selección, extracción de características y clasificación. El preprocesado consiste en una normalización (espacial y de intensidad) seguida de un enmascarado equivalente al de los atributos continuos para las ARs.

6.5.2 Selección de vóxeles mediante ARs

De la misma forma que se discutió en las secciones anteriores, dado que el patrón de imagen funcional para los controles muestra una variabilidad reducida cuando se compara con los patrones AD, el algoritmo *A priori* se usa para extraer ARs identificando relaciones relevantes entre regiones activadas de los sujetos normales. Las implicaciones ARs se establecen en términos de antecedentes y consecuentes entre regiones cerebrales simultáneamente activadas de controles. Las implicaciones AR se establecen en términos de sus valores de soporte y confianza que siempre están por encima de los umbrales de *minconf* y *minsup*. Además, las ARs se extraen en términos de la estrategia de validación cruzada LOO [Gidskehaug et al., 2008], es decir si un control determinado se está evaluando, no se tiene en cuenta en el proceso de extracción de ARs para de esta forma evitar el sobreentrenamiento.

La novedad de este método radica en el algoritmo de selección de vóxeles (véase sección 5.1 para detalles teóricos de los métodos de FS), que considera sólo los vóxeles que se involucran como antecedentes y consecuentes de las ARs extraídas en el paso previo. Los vóxeles se seleccionan al combinar sin repetición los antecedentes y consecuentes de las ARs previamente obtenidas que toman como valor el nivel de activación de esas ROIs (o suma de los puntos considerados como activos en esa ROI).

6.5.3 Extracción de características y clasificación

La reducción de la dimensionalidad (posterior a la FS) se consigue mediante técnicas de FE, en este contexto como el uso de PCA [López et al., 2011] o de PLS [Ramírez et al., 2010] tal como se motivó en el capítulo 5 de este trabajo. El conjunto resultante de características puede proveer una mejor habilidad de discriminación que el subconjunto de características dadas seleccionadas por las ARs. Por otra parte, la clasificación se lleva a cabo mediante un clasificador SVM que ha sido exitosamente usado en un gran número de aplicaciones [Adankon and Cheriet, 2009], [Noori et al., 2011], [Wang et al., 2011]. Además, los desarrollos recientes en la definición y entrenamiento de clasificadores estadísticos se hacen fiables en problemas de pequeño tamaño muestral e incluso son capaces de encontrar reglas de decisión no lineales para grupos de entrenamiento reducidos. Por último, se lleva a cabo la clasificación

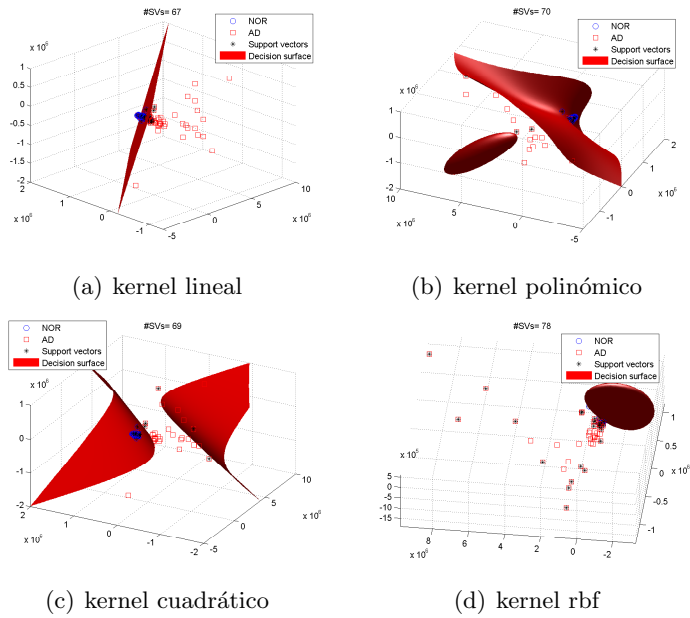


Figura 6.10: Superficies de separación para los diferentes kernels de SVM al usar el método de FS basado en ARs y de FE clásico de PCA. Base de datos SPECT.

mediante kernel SVM detallada en la sección 5.5.2. La figura 6.10 muestra las superficies de separación en SPECT para los diferentes kernels de SVM (en particular para el lineal, polinómico, cuadrático y rbf) cuando se aplica una FS mediante ARs y se reduce el espacio de características con PCA. En todos los casos, se manifiesta la habilidad de un clasificador SVM para separar las dos clases (controles en círculos frente a los pacientes AD en cuadrados) en término de los distintos kernels, aunque para la base de datos SPECT el clasificador lineal resulta más efectivo.

Capítulo 7

Vecinos más cercanos de amplio margen

En este capítulo, tras encuadrar el algoritmo LMNN en el campo de la métrica de distancias, introducimos su aplicabilidad a imagen médica funcional desde diferentes puntos de vista. En particular, LMNN puede ser usado como: i) una transformación del espacio de características para separar mejor las clases control y AD, ii) reductor de características mediante la transformación de una matriz rectangular (LMNN-RECT), y iii) como clasificador. La obtención de características se lleva a cabo en las fases de: enmascarado, cálculo del valor NMSE de los vóxeles tridimensionales centrados en las coordenadas de la máscara y selección de vóxeles mediante un t-test. En el caso i) la reducción del espacio de características se realiza mediante PCA/PLS seguida de la transformación LMNN, mientras que en ii) se aplica LMNN-RECT directamente y en iii) PCA/PLS. También se explica las bases del “truco del kernel” que simplifica la transformación de LMNN en kernel LMNN (KLMNN). Por último, la clasificación se efectúa mediante kernel SVM en los casos i) e ii) o bien haciendo uso del clasificador LMNN mediante las distancias Euclídeas, de Mahalanobis o los modelos basados en energía.

7.1 Introducción

La métrica de distancias es un asunto clave en los algoritmos de aprendizaje. Uno de los métodos más antiguos y simples para la clasificación de patrones es la regla de los k vecinos más cercanos (KNN, del inglés K nearest neighbors) [Wu et al., 2002]. La regla KNN clasifica cada ejemplo sin etiqueta por la mayoría de etiquetas de sus k vecinos más cercanos en el conjunto de entrenamiento. A pesar de su simplicidad, la regla kNN alcanza a menudo resultados competitivos [Weinberger et al., 2009]. KNN depende crucialmente de la forma en que se calculan las distancias entre ejemplos diferentes. Cuando no hay un conocimiento a priori disponible, la mayoría de las implementaciones de kNN calculan distancias euclídeas simples. El problema es que las distancias euclídeas ignoran cualquier regularidad estadística que pueda ser estimada mediante un conjunto grande de entrenamiento de ejemplos etiquetados. Se ha demostrado recientemente que una simple transformación lineal del espacio de características puede provocar mejoras significativas en la clasificación kNN [Goldberger et al., 2005], [Chopra et al., 2005]. Este capítulo profundiza en esta línea mostrando cómo aprender una métrica de distancias de Mahalanobis para la clasificación kNN [Chaves et al., 2011d], [Chaves et al., 2011b], [Chaves et al., 2011c]. Intuitivamente, este algoritmo (llamado LMNN, véase sección 5.4.1) se basa en la simple observación de que la regla de decisión kNN clasifica correctamente un ejemplo si sus k vecinos más próximos comparten la misma etiqueta. LMNN intenta incrementar el número de ejemplos de entrenamiento con esta propiedad mediante el aprendizaje de una transformación lineal del espacio de entrada. La transformación lineal se deriva por la minimización de una función de pérdidas como se explica teóricamente en la sección 5.4, que consta de dos términos. El primero penaliza las distancias grandes entre ejemplos de la misma clase, mientras que el segundo término penaliza las distancias pequeñas entre ejemplos con etiquetas que no se corresponden. Minimizar estos términos logra una transformación lineal del espacio de entrada que incrementa el número de ejemplos de entrenamiento cuyos k vecinos más cercanos tienen etiquetas que se corresponden. Las distancias euclídeas en el espacio transformado pueden verse como distancias de Mahalanobis en el espacio original.

7.2 Descripción del método

En primer lugar se realiza el mismo tipo de enmascarado que el descrito en la sección 6.2.1, es decir se tienen en cuenta las posiciones cuyo valor

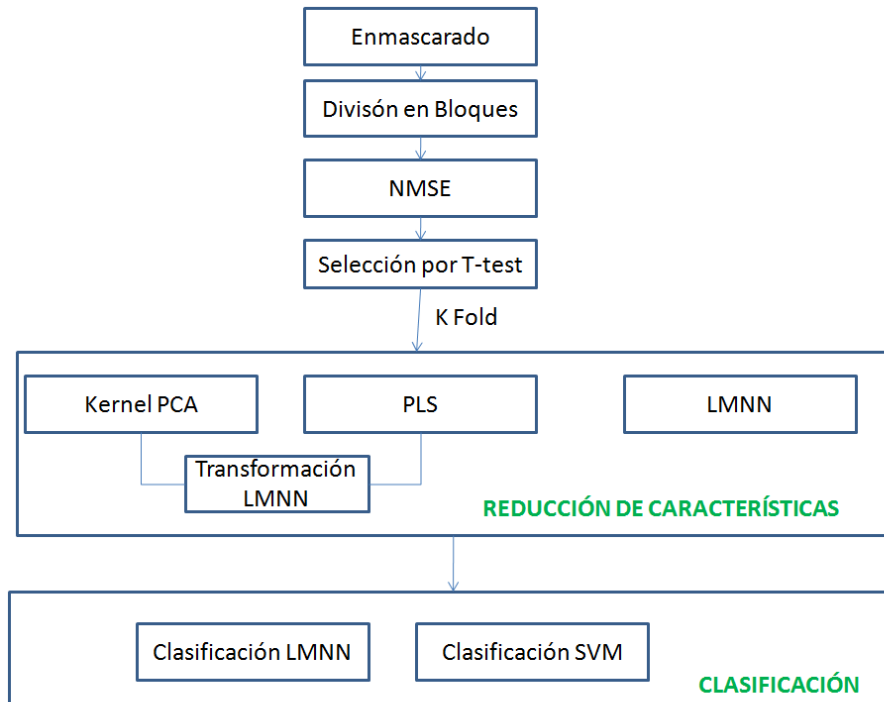


Figura 7.1: Extracción de características y diagrama de clasificación. Combinación de técnicas de reducción de características y clasificadores evaluados con validación cruzada de k -pliegues.

promedio de la intensidad de los normales se sitúe por encima de un umbral de activación a_T del 50%. En segundo lugar, se lleva a cabo la selección de características organizadas en bloques tridimensionales de tamaño $(2 \cdot v + 1) \times (2 \cdot v + 1) \times (2 \cdot v + 1)$. Además estos bloques adquieren el valor de su error cuadrático mínimo normalizado (NMSE) y se seleccionan con un t-test del mismo modo que en [Chaves et al., 2009b]. En tercer lugar, se reducen las características seleccionadas mediante PCA o PLS combinadas con una transformación LMNN para una mejor separación entre clases, o bien directamente mediante LMNN-RECT. Finalmente, la clasificación se lleva a cabo con clasificadores SVM o LMNN (basado en distancia euclídea, mahalanobis o con modelos de energía). Al clasificar con LMNN, este algoritmo no se utiliza en la extracción de características que se lleva a cabo sólo con PCA/PLS.

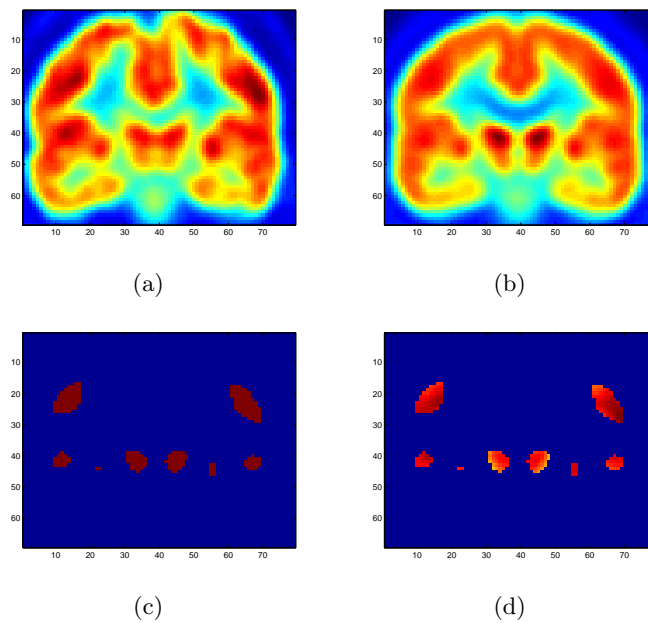


Figura 7.2: Corte coronal FDG-PET para mostrar el enmascarado en las fases de: a) original, b) promedio de controles, c) máscara, d) VAF: corte enmascarado

7.2.1 Extracción de características

En esta sección se propone aplicar una combinación de diferentes métodos de FE para obtener las características más importantes en el diagnóstico precoz de la AD. En ese sentido estamos reduciendo la complejidad del sistema al quitar las componentes inútiles y ruidosas. Además la FE resulta necesaria porque al enfrentarnos con el *problema de la maldición de la dimensión* las bases de datos presentan una alta dimensionalidad del espacio de características en comparación con el reducido número de sujetos [Chai et al., 2010]. Las diferentes fases de la extracción de características se encuentran representadas en la figura 7.1.

Enmascarado

En imagen funcional, cada vóxel muestra un nivel de intensidad de grises $I(x, y, z)$, que está relacionado con el flujo sanguíneo cerebral regional, el metabolismo de glucosa, etc. en el cerebro de un paciente, dependiendo de la modalidad de adquisición de la imagen. Los sujetos controles se promedian en una imagen tridimensional $\mathbf{f}(x, y, z)$ que se define como:

$$\mathbf{f}(x, y, z) = \frac{1}{C} \cdot \sum_{x=1}^X \sum_{c=1}^C f(x, y, z) + I_c(x, y, z) \quad (7.1)$$

donde C es el número total de controles. En segundo lugar se obtiene una máscara tridimensional $\mathbf{mask}(x, y, z)$ que consiste en todos los vóxeles que verifican:

$$\mathbf{f}(x, y, z) \geq a_T \quad (7.2)$$

donde el umbral a_T es equivalente al 50% de la máxima intensidad en $\mathbf{f}(x, y, z)$. Las figuras 6.2.a y 6.2.b muestran en el capítulo 6 las máscaras tridimensionales utilizadas para SPECT y FDG-PET (de ADNI) respectivamente.

VAF

En VAF [Stoeckel et al., 2001], las intensidades de los vóxeles en las coordenadas de la máscara obtenida, son directamente utilizadas para construir los vectores de características \mathbf{v} definidos como:

$$\mathbf{v} = (\mathbf{I}_1(\mathbf{mask}(x, y, z)), \dots, \mathbf{I}_P(\mathbf{mask}(x, y, z))) \quad (7.3)$$

donde P es el número total de pacientes de la base de datos. Por lo tanto, con el enmascarado los vóxeles fuera del cerebro y las regiones activadas pobremente se excluyen de este análisis. La figura 7.2 muestra un corte coronal de la base de datos FDG-PET con las fases descritas previamente: a) original, b) promedio de controles, c) máscara, d) VAF: corte enmascarado. El principal problema es la *maldición de la dimensionalidad*, dado que el número de muestras disponibles es mucho menor que el número de características usadas en el entrenamiento. Sin embargo en este capítulo, la combinación de técnicas de reducción de características no sólo resuelve este problema sino que también ayuda a alcanzar mejores resultados de clasificación.

Cálculo de NMSE en ROIs tridimensionales

En lugar de usar directamente todos los vóxeles, las ROIs se consideran en 3D porque no todas proveen el mismo valor discriminativo para detectar AD en su fase temprana. En efecto, el giro cingulado posterior y precuneus, así como la región temporoparietal se ven afectadas por la hypo-perfusión en AD [Chaves et al., 2009b]. Respecto a las regiones temporoparietales: muchos estudios han mostrado que esta región es típica para la AD; sin embargo, la mayoría de los estudios se llevaron a cabo con pacientes con un estado avanzado de AD que no se caracterizan por un desajuste cognitivo específico sino por una disminución cognitiva general. Así que esta región no describe AD precoz, aunque las anomalías temporoparietales bilaterales con o sin otros defectos regionales, se conocen como patrón predominante para la AD, no parecen ser ni cognitivas ni específicas para la AD temprana [Fung and Stoeckel, 2007]. Respecto al giro posterior cingulado y precuneiforme, los déficits de los niveles de perfusión son probablemente más específicos y frecuentes en AD precoz que los déficits temporoparietales [Kogure et al., 2000]. En lo que al lóbulo medio-temporal concierne: la hipoperfusión en estas regiones fue sólo dada a conocer durante el seguimiento a los pacientes. Algunas investigaciones sugieren que la hipoperfusión en estas regiones, como el hipocampo, no se encuentra en AD temprano debido a las dificultades de las imágenes, por lo que las herramientas que puedan asistir a los médicos para el estudio de estas regiones pueden ser muy beneficiosas.

En este capítulo, presentamos un método que no necesita explícitamente un conocimiento acerca del patrón de hipoperfusión de los pacientes AD. Por esa razón cada imagen funcional se procesa por medio de vóxeles cúbicos tridimensionales de tamaño $(2 \cdot v + 1) \times (2 \cdot v + 1) \times (2 \cdot v + 1)$ definiendo ROIs centradas en las coordenadas de la máscara. A continuación, se calcula el Error Mínimo Cuadrático Medio o $NMSE_p(x,y,z)$ [Chaves et al., 2009a]

definido como:

$$NMSE_p(x, y, z) = \frac{\sum_{l,m,n=-v}^v [f(x-l, y-m, z-n) - I_p(x-l, y-m, z-n)]^2}{\sum_{l,m,n=-v}^v [f(x-l, y-m, z-n)]^2} \quad (7.4)$$

que se obtiene para cada sujeto y bloque (véase figura 7.1) donde $f(x, y, z)$ es la intensidad media del vóxel de todos los sujetos controles y $I_p(x, y, z)$ es la intensidad de vóxel de sujeto p -ésimo en las coordenadas (x, y, z) . Las ROIs más discriminativas se obtienen por medio de un valor absoluto de dos muestras t -test con estimación de la covarianza combinada en las características NMSE como se explica en [Chaves et al., 2009b].

Selección de características con t-test

El t-test [Schechtman and Sherman, 2007] está basado en un t-test absoluto de dos muestras con estimación de varianza acumulada en características NMSE:

$$T(x, y, z) = \frac{|\mu_1 - \mu_0|}{\sqrt{\sigma_1^2 + \sigma_0^2}} \quad (7.5)$$

donde μ_1 y μ_0 son las medias de clase AD y control, y σ_1^2 y σ_0^2 sus varianzas respectivamente. La selección de características también utiliza información de correlación para obtener el valor T de las características potenciales $T \cdot (1 - \alpha \cdot \rho)$ donde ρ es el promedio de los valores absolutos del coeficiente de correlación cruzada entre la característica candidata y todas las características previamente seleccionadas y α es un valor escalar entre $[0, 1]$ que proporciona el factor de pesado. Cuando α es 0, las características potenciales no se pesan y cuando α es cercano a 1, significa que las características se correlacionan fuertemente con las características ya elegidas y son menos probables de ser incluidas en la lista de salida. En los experimentos mostrados en este trabajo, es 0.95. La figura 7.3 muestra un mapa del valor T (valor discriminativo de la característica) para cada voxel dentro de la máscara con un tamaño de vóxel $5 \times 5 \times 5$.

Reducción de características

Como se explicó en la sección 5.3 y se especifica en la figura 7.1, una gran variedad de métodos utilizados para la extracción de características son PCA [Jackson, 1991], [Wold et al., 1987] y las proyecciones a estructuras latentes (PLS) [Wold, 1982], [Tenenhaus, 1998] que funcionan computacionalmente

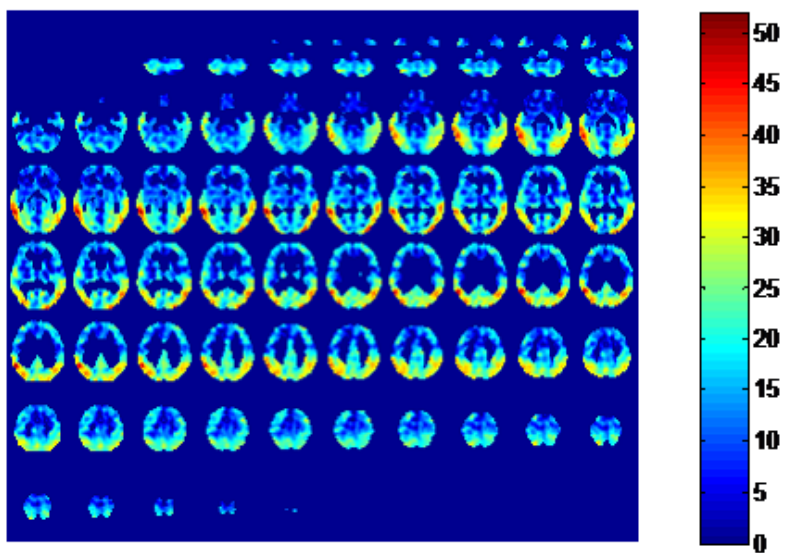


Figura 7.3: Magnitud de T para cada uno de los vóxeles dentro de la máscara y con un tamaño $5 \times 5 \times 5$

bien para muchas variables y observaciones. Por otro lado, el algoritmo LMNN tiene como objetivo acercar los k vecinos más cercanos de la misma clase, mientras que los de diferentes clases se separan por un margen grande [Weinberger et al., 2009], [Goldberger et al., 2005], [Chopra et al., 2005], [Tsochantaridis et al., 2005]. Como se detalla en la sección 5.4.2 (para la transformación LMNN y LMNN-RECT), LMNN puede ser usado como transformador del espacio de características, en este caso obtenidas con técnicas de extracción como PCA o PLS (aplicado sobre ROIs preseleccionadas con NMSE y ttest); o bien directamente como técnica de extracción de características (LMNN-RECT) de las ROIs. En este trabajo proponemos y comparamos estos métodos de extracción de características (mostrados en la figura 7.1).

Truco del Kernel

En muchos casos, necesitamos considerar transformaciones no-lineales de datos para aplicar los algoritmos de aprendizaje. Un método eficiente para ello es utilizar un núcleo (o “kernel” en inglés) que calcule una medida de similitud entre otros dos puntos. En este capítulo mostramos cómo podemos aprender una métrica de distancias en el espacio de características mediante un “kernel”, para realizar mapeados no-lineales dado que en muchos problemas de aprendizaje de distancias, las distancias euclídeas entre puntos no representan la topología que intentamos mostrar. Así los métodos de “kernel” mapean los puntos a nuevos espacios donde las distancias euclídeas pueden ser más útiles. Conceptualmente, estamos mapeando los puntos en un espacio de características de forma no-lineal $\phi()$ y posteriormente aprendiendo una métrica de distancias en ese espacio. De hecho, llevar a cabo este mapeado típicamente no es deseable, así que se emplea el conocido “truco del kernel” $K(x_i, x_j)$ que puede calcular productos internos entre los vectores de características sin construirlos explícitamente [Alipanahi et al., 2008]. En este capítulo, el truco del kernel se utiliza para transformar LMNN, en kernel LMNN (KLMNN). Ello se consigue al combinar PCA o PLS kernelizados (mediante kernel gaussiano) con LMNN. Teóricamente las distancias cuadradas en nuestro objetivo tienen la forma:

$$(x_i - x_j)^T \cdot W \cdot W^T \cdot (x_i - x_j) \quad (7.6)$$

Esta matriz W puede expresarse como una combinación lineal de los puntos $W = X \cdot \beta$, mediante el truco del kernel. Al reescribir la distancia cuadrada queda:

$$(k_i - k_j)^T \cdot \beta \cdot \beta^T \cdot (k_i - k_j) \quad (7.7)$$

donde $k_i = X^T \cdot x_i$ es la i -ésima columna de $K = X^T \cdot X$. Hemos expresado la distancia en términos de productos internos entre los puntos de los datos, que pueden ser calculados vía el kernel K .

7.2.2 Clasificación

Los clasificadores utilizados en esta sección son LMNN (usando distancia Euclídea, Mahalanobis o con métodos basados en energía) y kernel SVM como se explica en la sección 5.5. LMNN como reductor o transformador del espacio de características se combina con kernel SVM, mientras que no con los clasificadores basados en LMNN. El método de evaluación utilizado es k -pliegues que se explica en la sección de fundamentos teóricos 5.6. LMNN puede verse como el homólogo lógico de SVM con algunas diferencias descritas en la sección 5.5.4. La metodología de amplio margen consigue intuitivamente la mayor robustez frente al ruido y los outliers. Ese riesgo se mitiga en SVM normalmente utilizando un criterio de margen suave, de la misma forma que la función de pérdidas de LMNN es utilizada para reducir la sensibilidad de los outliers. La figura 7.4 muestra las superficies de separación para los diferentes kernels de SVM (en particular para el lineal, polinómico, cuadrático y rbf) cuando se aplica PCA en combinación con la transformación LMNN para la extracción de características siguiendo el método descrito en este capítulo. Además la figura 7.5 muestra las superficies de separación para los diferentes kernels cuando la extracción se lleva a cabo con LMNN-RECT. En ambos casos (con la transformación LMNN o LMNN-RECT), se pone de manifiesto la habilidad de un clasificador SVM para separar las dos clases (controles en círculos frente a los pacientes AD en cuadrados) en término de los distintos kernels.

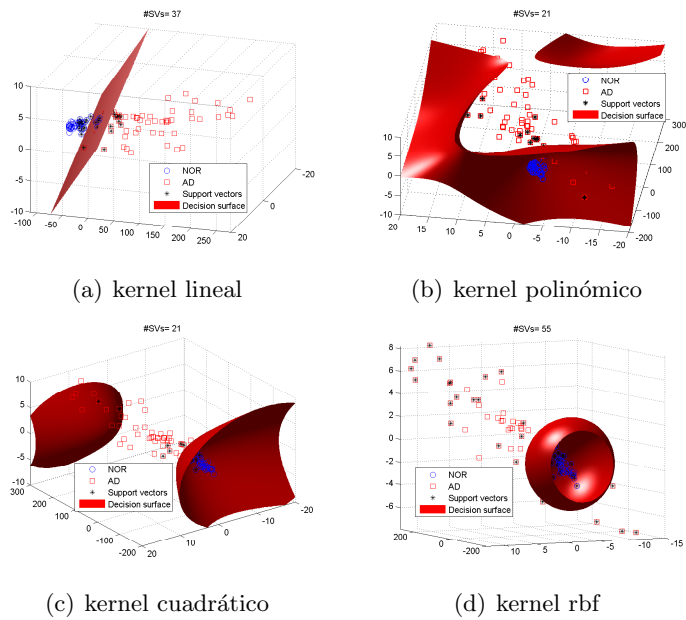
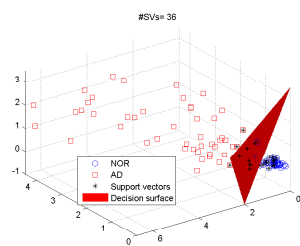
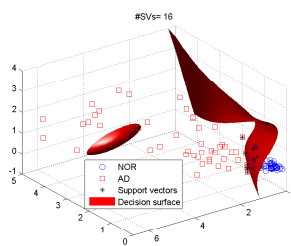


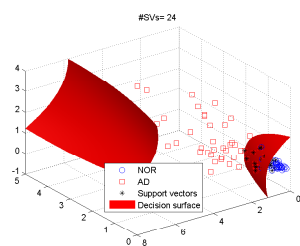
Figura 7.4: Superficies de separación para los diferentes “kernels” de SVM al usar el método de transformación LMNN junto con PCA como extractor de características. Base de datos SPECT.



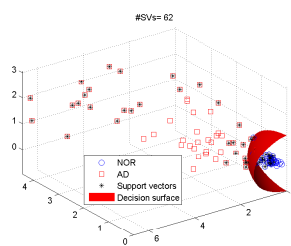
(a) kernel lineal



(b) kernel polinómico



(c) kernel cuadrático



(d) kernel rbf

Figura 7.5: Superficies de separación para los diferentes “kernels” de SVM al usar el método de reducción de características LMNN-RECT. Base de datos SPECT.

Parte V

Resultados experimentales

Capítulo 8

Bases de datos utilizadas

Las bases de datos de neuroimagen médica funcional utilizadas en la sección experimental (capítulos 9 y 10) de este trabajo facilitan la comprobación de la validez y la robustez de los métodos propuestos en esta Tesis y constan de: una base de datos SPECT procedente del Hospital Universitario Virgen de las Nieves en Granada, una base de datos FDG-PET del Hospital Cartuja de Sevilla, por otro lado desde la organización ADNI para modelar la AD y sus progresión, empleamos una base FDG-PET de 150 pacientes así como una base de datos de 103 pacientes adquirida por separado mediante los biomarcadores FDG y PIB para investigar la complementariedad de los mismos. Los detalles demográficos así como los proporcionados por los tests neuropsicológicos, las características de la normalización empleada o las regiones cerebrales afectadas por la AD son factores que se han tenido en cuenta a la hora de describir las bases de datos presentadas.



Figura 8.1: Cámara gamma para la obtención de SPECT del Hospital Universitario Virgen de las Nieves, Granada.

8.1 SPECT

Las imágenes SPECT usadas en este trabajo fueron adquiridas en términos de un cámara gamma (como la de la figura 8.1 procedente del Hospital Universitario Virgen de las Nieves, Granada) después de inyectar tecnecio-99m etiquetado dímero etílicocisteinato ($^{99m}\text{Tc-ECD}$) a cada sujeto. Estas imágenes fueron reconstruidas a partir de las proyecciones obtenidas mediante retroproyección filtrada (FBP) en combinación con un filtro de Butterworth para eliminar ruidos tal como se explica en la sección 3.4. Las imágenes SPECT se normalizaron [Toga, 1999], [Salas-Gonzalez et al., 2008] con posterioridad espacialmente para asegurar que un vóxel dado en diferentes imágenes se refiere a la misma posición anatómica. Este proceso se realizó mediante mapas estadísticos paramétricos (SPM) [Friston et al., 2007] consiguiendo imágenes SPECT normalizadas de tamaño $69 \times 95 \times 79$. Una comparación directa de las intensidades de las imágenes de los diferentes sujetos no es posible sin la nor-

Detalles demográficos			
	Num. de muestras	Sexo (M/F) (%)	Edad μ [rango/ σ]
CTRL	41	32.95/12.19	71.51 [46-85/7.99]
AD 1	30	10.97/18.29	65.20 [23-81/13.36]
AD 2	22	13.41/9.76	65.73 [46-86/8.25]
AD 3	4	0/2.43	76[69-83/9.90]

Tabla 8.1: Información demográfica de la base de datos SPECT. μ and σ representa la media de la población y la desviación estándar respectivamente.

malización de intensidades [Saxena et al., 1998]. El nivel de intensidad de las imágenes se normaliza a la máxima intensidad, que se realiza mediante ordenador para cada imagen individualmente promediando sobre el 0.1% de la intensidad más alta del nivel de voxel [López et al., 2009a]. El proceso de normalización (espacial y en intensidad) se explica en este trabajo en la sección 3.4.4.

La base de datos consta de estudios de imagen de los sujetos siguiendo el protocolo de un hospital. En primer lugar, el neurólogo evalúa la función cognitiva, y aquellos pacientes a los que se le ha detectado pérdidas de memoria o demencia se les remite al departamento de medicina nuclear en el Hospital Virgen de las Nieves de Granada, para adquirir información complementaria de imagen para el diagnóstico. Los médicos expertos en el campo evalúan las imágenes visualmente. Las imágenes se evalúan usando 4 etiquetas diferentes: Controles (CTRL) para sujetos sin anormalidades gammagráficas y déficit de perfusión temprana (AD1), déficit moderado (AD2) y déficit severo (AD3), para distinguir entre diferentes niveles de presencia de patrones de hipoperfusión compatibles con AD. La base de datos consta en total de 97 sujetos: 41 CTRL, 30 AD1, 22 AD2 y 4 AD3 (ver la tabla 8.1 para detalles demográficos). Dado que los pacientes no están confirmados patológicamente, las etiquetas de los sujetos poseen cierto grado de incertidumbre, ya que el patrón de hipoperfusión puede no reflejar la patología subyacente de AD, ni tampoco que la clasificación de los scans necesariamente refleja la severidad de los síntomas de los pacientes. Sin embargo, cuando la información patológica está disponible, la evaluación visual por expertos se ha mostrado muy sensible y específica como método de etiquetado, en comparación con los tests neuropsicológicos [Jobst et al., 1998], [Dubois et al., 2007]. Dado que esta es una limitación de los estudios *in vivo*, nuestra suposición de trabajo es que las etiquetas son ciertas, considerando la etiqueta del sujeto positiva cuando pertenece a cualquiera de las clases de AD, y negativa en otro caso. Este trabajo no implica ninguna

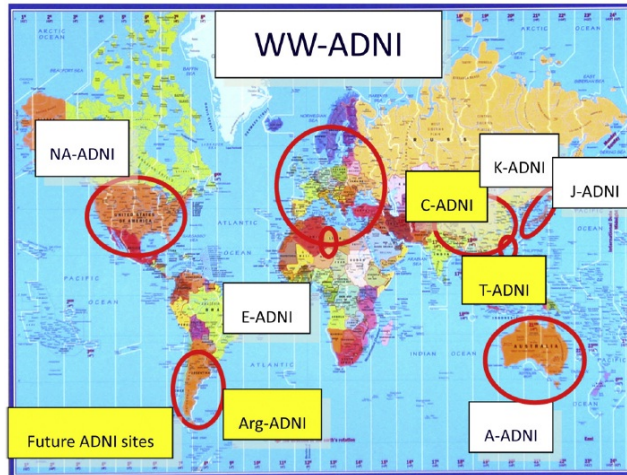


Figura 8.2: ADNI a nivel mundial: NA-ADNI (Norte América), Arg-ADNI (Argentina), E-ADNI (Europea), C-ADNI (China), K-ADNI (Coreana), J-ADNI (Japonesa), T-ADNI (Taiwanesa), A-ADNI (Australiana). Figura obtenida de [Weiner et al., 2012].

intervención experimental y ha sido realizado bajo la aprobación y supervisión de la comisión ética médica y de investigación (CEIC) del Hospital Universitario Virgen de las Nieves. ¹.

8.2 ADNI-PET

Las imágenes PET (explicadas teóricamente en la sección 3.3.2) se obtuvieron del laboratorio ADNI de Neuroimagen (LONI, Universidad de California, Los Angeles) con página web (<http://www.loni.ucla/ADNI/>). La ADNI fue fundada en 2003 por el Instituto Nacional del Envejecimiento (National Institute on Aging, NIA), el Instituto Nacional de Imagen Biomédica y de Bioingeniería (National Institute of Biomedical Imaging and Bioengineering, NIBIB), la administración de alimentos y medicinas (Food and Drug Administration, FDA), compañías privadas farmacéuticas y organizaciones sin ánimo de lucro como socio público-privado de 60 millones de dólares y 5 años.

Desde la creación de ADNI en América del Norte, ha existido un interés mundial para la creación de programas que estén al menos parcialmente mode-

¹La información clínica no está disponible por razones de privacidad, sólo se dispone de la información demográfica

lados en la plataforma ADNI, y que usen sus protocolos durante al menos parte de sus estudios. El beneficio puede radicar no sólo en tener un seguimiento de la sociedad norteamericana para el desarrollo de los biomarcadores AD sino también a nivel global para diversos grupos étnicos, importantes para determinar la aplicación de los descubrimientos ADNI a la población mundial. Como ADNI, estas iniciativas desde Europa, Japón y Australia predicen el compartir datos, y la infraestructura está comenzando a ser desarrollada para permitir la transparencia global. El futuro de ADNI está orientado a países como Argentina y China y ha comenzado recientemente en Corea y Taiwan [Weiner et al., 2012]. Todos los subgrupos ADNI comparten objetivos comunes tales como diagnosticar precozmente el AD y la comprensión, tanto cognitiva como física, estableciendo estándares globalmente reconocidos para el diagnóstico, y últimamente desarrollando métodos para permitir procesos clínicos más eficientes. Los lugares donde ADNI trabaja a nivel mundial aparecen representados en el mapa de la figura 8.2.

El objetivo prioritario de ADNI ha sido testear si la imagen de resonancia magnética (MRI), PET, otros marcadores biológicos, o la evaluación clínica y neuropsicológica puede ser combinada para medir la progresión del desajuste cognitivo temprano (MCI) y AD temprano. Se pretende que la determinación de los marcadores sensibles y específicos de la progresión del AD en sus estadios más tempranos pueda ayudar a los investigadores y médicos a desarrollar nuevos tratamientos y monitorizar su efectividad, así como aliviar el tiempo y coste de los ensayos clínicos [Shin et al., 2010].

Existen numerosos estudios centrados en las relaciones entre la función cognitiva, de imagen o de biomarcadores CSF [Hampel et al., 2003] (ver secciones 2.4 y 2.5 para más información). Los cambios de atrofia en los episodios de pérdida de memoria se centran en las estructuras MTL (lóbulos medio-temporales), áreas corticales parietales medias y laterales, y áreas corticales prefrontales involucradas en la evolución de la pérdida de memoria de la AD como se muestra en la figura 8.3. Los estudios ADNI concluyen en el seguimiento de la base de datos que el volumen del hipocampo y el metabolismo, grosor del parahipocampo y genotipo APOE predicen el reconocimiento; mientras que el volumen y metabolismo del hipocampo, el grosor cortical del precuneus y el metabolismo parietal inferior predicen el aprendizaje. Esto sugiere que las estructuras medio temporales se relacionan con el aprendizaje, memoria y reconocimiento, mientras que las estructuras parietales se involucran únicamente en el aprendizaje [Weiner et al., 2012]. La figura 8.3 muestra las regiones donde se detecta la evolución de la pérdida de memoria entre las que destacan las regiones temporoparietales, el giro poste-

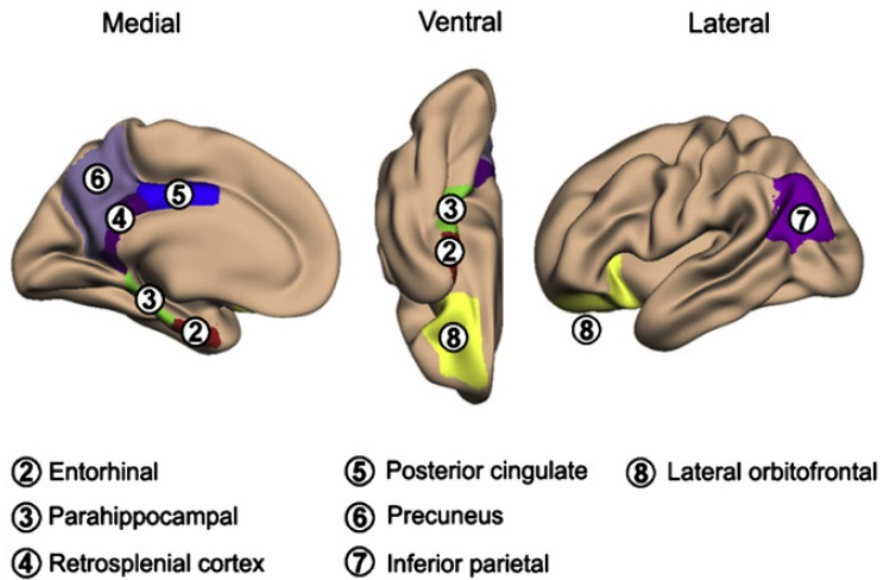


Figura 8.3: Regiones donde se detecta la evolución de la pérdida de memoria [Weiner et al., 2012]

rior cingulado y precuneiforme, el hipocampo, la corteza entorrinal etc.

El investigador principal de esta iniciativa es Michael W. Weiner, MD, del centro médico VA y la Universidad de California (San Francisco). ADNI es el resultado de esfuerzos de muchos co-investigadores de un amplio rango de instituciones académicas y corporativas privadas, y sujetos que han sido reclutados de 50 sitios de los Estados Unidos y Canadá. El objetivo inicial de ADNI fue la búsqueda de 800 adultos, de edades entre 50 a 90 años (dado que es donde la AD tiene mayor prevalencia), para participar en la investigación, aproximadamente 200 pacientes cognitivamente normales individualmente seguidos durante 3 años, 400 personas con MCI para ser seguidos durante 3 años y 200 personas con AD en fase temprana seguidos durante 2 años. Para información actualizada, ver www.adni-info.org.

8.2.1 ^{18}F -FDG ADNI PET

El metabolismo de la glucosa regional cerebral se estima típicamente usando ^{18}F -FDG PET, es decir, la acumulación de este biomarcador de imagen [Bonmati et al., 2012] en una región cerebral dada es proporcional a su ratio metabólico de glucosa. En los casos típicos de AD, el metabolismo de glu-

cosa disminuye (hipometabolismo) [Benoit et al., 2002] y el fluido sanguíneo cerebral inicialmente se encuentra en el corte posterior parieto-temporal y en el corte posterior cingulado. Durante la evolución de AD, el declive metabólico se reparte progresivamente. El hipometabolismo revelado por PET en los pacientes AD refleja varios procesos patológicos que subyacen tales como la pérdida de la actividad sináptica, gliosis y los depósitos de placas de amiloide. Otros factores de vulnerabilidad para la AD son la presencia de polimorfismos genéticos para la apolipoproteína E epsilon (APOE). Varios estudios concluyen que PET es sensible de mostrar hipometabolismo años antes de que se manifiesten los síntomas de la AD [Rosa-Neto and Leuzy, 2009].

Las imágenes FDG PET se adquirieron de acuerdo a un protocolo estandarizado como se explica en la sección 3.3.4. Una imagen de emisión dinámica, consta de imágenes adquiridas 30 minutos después de la inyección intravenosa de 5.0 ± 0.5 mCi de ^{18}F -FDG. Los pacientes tuvieron que ayunar 4 horas previamente al scan, se quedaron en reposo en una habitación poco iluminada con los ojos abiertos y estimulación sensorial mínima. La atenuación de radiación y dispersión de los datos fue corregido usando escáneres de transmisión de fuentes Ge-68 rotativos de barra y reconstruidos usando corrección de atenuación y algoritmos de reconstrucción de imágenes especificadas para cada escáner (ver más detalles teóricos de reconstrucción en la sección 3.4). Seguidamente cada imagen fue revisada en la Universidad de Michigan para mitigar los posibles artefactos y se archivaron todos los datos de estudio raw y procesados.

A continuación, las imágenes fueron normalizadas a través de un modelo general afín, con 12 parámetros [Salas-Gonzalez et al., 2008] usando el software SPM5. Después de la normalización afín, la imagen resultante fue registrada usando un modelo de transformación espacial no rígida. Las deformaciones no lineales de la Plantilla de Imagen Neurológica de Montreal (Montreal Neurological Imaging, MNI) fueron parametrizadas por una combinación lineal de las componentes de frecuencia más bajas de las bases de transformación coseno [Ashburner and Friston, 1999]. Se usa una aproximación de deformación pequeña, y la regularización de la energía concentrada del campo de desplazamiento, asegurando que los vóxeles en imágenes diferentes FDG-PET se refieran a la misma posición anatómica en el cerebro. Después de la normalización espacial, una normalización de intensidad se necesita para poder realizar comparaciones directas entre las imágenes de sujetos. La intensidad de las imágenes se normalizó a un valor de intensidad I_{max} , que se obtuvo promediando el 0.1% de las intensidades de los vóxeles más altas que exceden un umbral con un método similar al expuesto en [Saxena et al., 1998]. El umbral

Detalles demográficos			
	Núm. de muestras	Sexo (M/F) (%)	Edad μ [rango/ σ]
CTRL	75	58.67/41.33	75.97 [62-86/4.91]
AD	75	62.67/37.33	75.72 [55-88/7.40]

Tabla 8.2: Detalles demográficos de la base de datos FDG-PET: μ y σ se refieren a la media y la desviación estándar de la población respectivamente

fue fijado al valor del décimo bin de intensidad de un total de un histograma de 50 bins, para descartar la mayoría de registros de baja intensidad de regiones fuera del cerebro y prevenir la saturación de imágenes. La selección de participantes se condicionó a criterios generales de inclusion-exclusión:

- Sujetos normales (CTRL): valores MMSE entre 24–30 (incluidos), CDR de 0, no deprimido, no MCI, y no dementes. El ratio de edad de los sujetos normales se asocian rara vez a la de los sujetos MCI y AD, que suele situarse por encima de los 70 años.
- AD precoz: valores MMSE entre 20–26 incluidas, CDR de 0.5 o 1.0, y satisface el criterio NINCDS/ADRDA para AD probable.

La base de datos PET de ADNI utilizada en este trabajo consta de 150 imágenes PET: 75 controles y 75 AD (véase la tabla 8.2 para detalles demográficos). Los diagnósticos de los pacientes ADNI no están patológicamente confirmados, introduciendo alguna incertidumbre en las etiquetas. Usando estas etiquetas, se puede probar la robustez del clasificador. Esto debe ser también considerado cuando se compara con otros métodos testeados en pacientes AD de autopsia confirmada, de los que se espera que cada clasificador funcione mejor [Illán et al., 2011a].

8.2.2 ^{11}C -PIB PET

Un método complementario para evaluar el depósito de amiloide ($A\beta$) es la imagen ^{11}C -PIB [Shin et al., 2010]. En comparación con los controles, los pacientes con AD muestran la retención de PIB fundamentalmente en el corte frontal, con grandes incrementos en los cortes parietal, temporal y occipital y en el estriatum. Sin embargo, y al contrario que con FDG, la retención de PIB es baja en las áreas corticales [Klunk and et al., 2004]. Se ha demostrado que existe un acuerdo sustancioso entre PET CSF $A\beta$ -42 y ^{11}C -PIB en lo que a medidas del depósito de amiloide se refiere aunque no para los valores del test MMSE [Brugnolo et al., 2009]. En contraste, FDG-PET como medida

del metabolismo de la glucosa se relaciona mejor con las puntuaciones MMSE, pero peor con los biomarcadores CSF. Por otro lado, aunque todos los biomarcadores CSF se asocian con la atrofia del hipocampo, las correlaciones mayores se producen con $p - tau_{181p}$ y la más débil con $A\beta-42$ [Weiner et al., 2012]. ^{11}C -PIB precuneal está fuertemente asociado con la atrofia del hipocampo. En la progresión desde MCI a AD fundamentalmente las atrofas del hipocampo y el depósito de amiloide, se miden como una composición de ^{11}C -PIB y $A\beta-42$ CSF [Weiner et al., 2012]. Adquiridos desde diferentes escáneres, 103 participantes fueron elegidos por el laboratorio LONI ADNI condicionados a algunos criterios de selección. Los criterios generales de inclusión y exclusión fueron los siguientes [Illán et al., 2010]:

- CTRL: valores MMSE entre 24-30 incluidos, con un CDR de 0 que significa no deprimido, no MCI y no demente. El rango de edad de los sujetos normales estará bastante relacionado con el de los sujetos MCI y AD. Por lo tanto, debe existir una selección mínima de sujetos normales más jóvenes de 70 años.
- MCI: los valores MMSE están entre 24-30 incluidos, una queja de pérdida de memoria medida por la escala de memoria lógica de Wechsler II, un CDR de 0.5, que simboliza la ausencia de niveles significativos de desajuste en otros niveles cognitivos, esencialmente en actividades de la vida diaria y en una ausencia de demencia.
- AD temprano: puntuaciones MMSE entre 20-26 incluidos, CDR de 0.5 o 1.0, y cumple el criterio NINCDS/ADRDA para AD probable.

De esta forma PET está compuesto de 3 clases: 19 CTRL (rango de edad en media \pm desviación estándar: 76.5 ± 4.8), 65 MCI (75.1 ± 7.4) y 19 AD (77.2 ± 7.2). Un conjunto preprocesado está disponible de la base de datos ADNI, en la que cada imagen del sujeto PET se corregistra con la línea AC-PC, promediada, estandarizada espacialmente y en intensidad y con resolución uniforme. Estos pasos tratan de minimizar las diferencias procedentes del uso de escáneres diferentes, diferentes resoluciones, movimientos durante la adquisición etc. Para la comparación a nivel de vóxel entre sujetos, la base de datos fue normalizada espacialmente usando el software SPM5. Esto se consiguió utilizando el mismo tipo de normalización que la explicada para la base de datos FDG-PET en la sección 3.4.4, es decir con una normalización a la plantilla MNI y un modelo general afín con 12 parámetros y una función de coste. Uno de estos parámetros se obtuvo de la imagen FDG, y cada imagen PIB corregistrada se normalizó usándolos, garantizando exactamente el mismo

procesado para ambas modalidades. En este trabajo para poder comparar diferentes marcadores (PIB y FDG) y su combinación, se seleccionaron los mismos 103 pacientes pero analizados separadamente con cada biomarcador de imagen.

8.2.3 Modelo de la enfermedad y progresión utilizando ADNI

El modelo empieza con el depósito anormal de fibrillas $A\beta$ [Findeis, 2007], como se evidencia en la caída de los niveles solubles del $A\beta$ -42 en el CSF [Hampel et al., 2003] y el incremento de la retención del trazador radiactivo de PET [^{11}C]- etiquetada como componente B de Pittsburg ^{11}C -PiB en el cortex cerebral. Algo más tarde el daño neuronal comienza a producirse como evidencia del incremento de los niveles de la proteína CSF tau. La disfunción sináptica continúa produciendo un decrecimiento de la absorción de [^{18}F]-FDG medida por PET [Weiner et al., 2012]. En lo que a la progresión de la degeneración neuronal se refiere, la atrofia en ciertas áreas típicas de AD es detectada por la imagen de resonancia magnética (MRI). Respecto a los estudios de seguimiento de ADNI, un ejemplo de modelo que propone una serie de eventos patológicos dirigidos hacia el desajuste cognitivo y la demencia aparece en la figura 8.4. La figura 8.4 muestra la evolución de la fase clínica de la enfermedad (desde normal, desajuste cognitivo temprano hasta demencia) frente a la magnitud del biomarcador que muestra la progresión desde la normalidad a la anormalidad cuando se comparan los marcadores $A\beta$, tau, de estructura cerebral, memoria o función clínica. Se deduce de esta figura que la detección temprana de AD se realiza con mayor facilidad cuando se utilizan los biomarcadores $A\beta$ o tau, detectando la demencia en estados bastante más avanzados que al hacer uso de los marcadores de estructura cerebral, memoria y función clínica. Esta figura está relacionada con la bondad de la imagen médica (que refleja el estadio de los biomarcadores) para detectar precozmente el AD frente a otros marcadores.

8.3 ^{18}F -FDG PET Cartuja

La base de datos PET consiste en 60 imágenes (18 normales y 42 sujetos AD en estado severo) provistos por la clínica privada Cartuja (Sevilla, España) lo que supone una prevalencia de alrededor del 70% ya que ésta cuantifica la proporción de individuos de la clase AD (en un momento o período de tiempo determinado) que ha de tenerse en cuenta en análisis posteriores. Las etiquetas de las imágenes se asignaron por un grupo de clínicos expertos y se asume como *patrón oro* dado que saber con 100% de seguridad si un sujeto

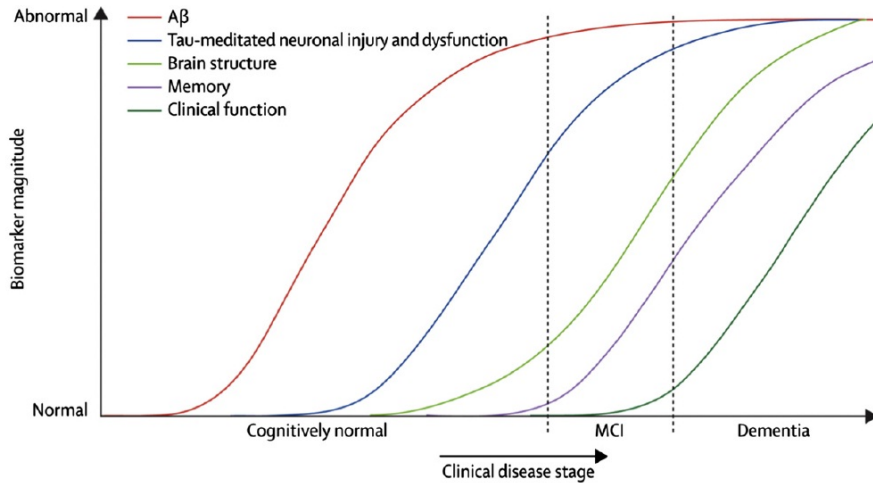


Figura 8.4: Modelo para la progresión de AD. Obtenida de [Weiner et al., 2012]

sufre la enfermedad es imposible hasta después de un análisis post-mortem [Segovia et al., 2010].

El radiofármaco utilizado es el ^{18}F -FDG y la cámara fue la Siemens CTI ECAT EXACT 47. La adquisición de imágenes se produce 30 minutos (es importante fijar un tiempo para realizar comparativas fiables entre pacientes) desde la administración del radiofármaco ^{18}F -FDG al paciente, tiempo durante el cual el paciente permanece en reposo en una habitación en silencio y con iluminación tenue para que el radiofármaco se distribuya correctamente. Para la adquisición, el paciente se posiciona de cúbito supino con la cabeza fijada (con una cinta por ejemplo) para evitar movimientos durante los 30 minutos en los que la cámara PET recoge las señales emitidas por el contraste y las convierte en imágenes funcionales [Pérez, 2010].

El paso de posfiltrado consiste en un filtro gaussiano y en un filtro axial. Para asegurar que se analizan las mismas posiciones en el sistema de coordenadas, todas las imágenes han sido normalizadas espacialmente con SPM logrando un mapa de actividad funcional tridimensional $69 \times 95 \times 79$ para cada paciente. Además, la intensidad de los vóxeles se normalizó ecualizando el valor máximo y escalando las restantes.

Capítulo 9

Resultados experimentales con ARs

Motivado por el capítulo 6 de análisis de imagen funcional mediante ARs, presentamos a continuación los correspondientes resultados experimentales organizados en: ARs entre atributos continuos, discriminantes o discretos. Como novedad las ventajas de la combinación de diversos biomarcadores proporcionados por estudios PET con los biomarcadores ^{11}C -PIB y ^{18}F -FDG se muestran para los atributos continuos. También se presentan los resultados cuando se utilizan atributos discriminantes obtenidos mediante el umbral FDR para mejorar el gasto computacional. Los atributos discretos son una alternativa para evitar el uso de umbrales y automatizar el proceso experimental. Por último presentamos las ARs como selector de características condicionados a parámetros como la activación de las ROIs o los umbrales minsup y minconf del algoritmo Apriori para la extracción de ARs. Todas las secciones fueron analizadas con distintas bases de datos (descritas en el capítulo 8) para resaltar la generalidad y la robustez de los métodos propuestos.

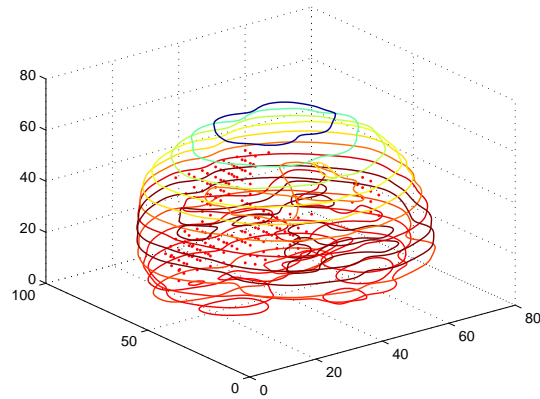


Figura 9.1: SPECT. ROIs tridimensionales (centradas en los puntos) para un tamaño de voxel de 9, espacio de grid $g=4$ y activación por bloque de 90%

9.1 ARs entre atributos continuos

En esta sección, mostramos los resultados experimentales tras aplicar los métodos propuestos en la sección 6.2 para ARs entre atributos continuos. Los experimentos se evalúan con bases de datos SPECT, PET Cartuja y de biomarcadores (11C-PIB y 18F-FDG). En esta última, el hecho de combinar biomarcadores de imagen como FDG y PIB favorece los resultados de clasificación, dado que se combina la información que ambos aportan, enriqueciendo los resultados obtenidos. Se estudian los conjuntos controles frente a los de AD en PET Cartuja y SPECT, para diagnosticar tempranamente la enfermedad. En la base de datos de biomarcadores, se considera también el grupo MCI frente al de control en la evaluación dado que la combinación de FDG y PIB mejora considerablemente los resultados.

El aprendizaje supervisado propuesto basado en ARs se aplica a un número reducido de ROIs consiguiendo exactitudes de clasificación mejores que con las aproximaciones VAF [Stoeckel et al., 2004] o PCA [Friston et al., 2007], [Álvarez et al., 2009] que se utilizan como comparativa. La reproducción del procedimiento visual logrado por los médicos para la identificación de patrones de hipoperfusión se consigue de una forma más dinámica que con otros métodos, dado que las ARs se basan en las relaciones establecidas entre regiones cerebrales que por su activación han sido seleccionadas como interesantes.

Como se explicó en el capítulo 6, algunos experimentos se realizaron para

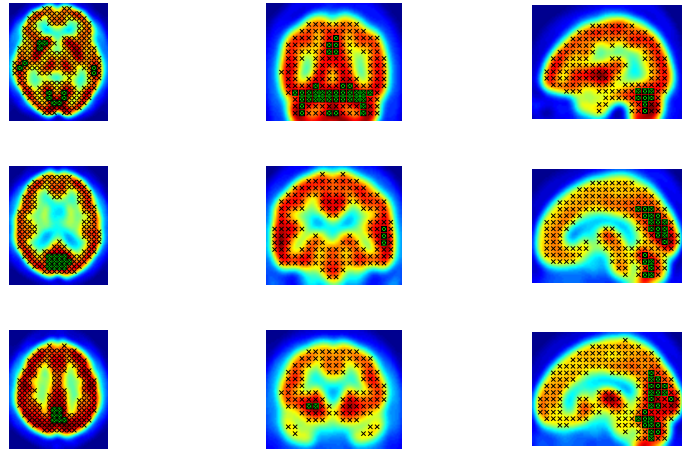


Figura 9.2: SPECT, método AR entre atributos continuos. ROIs activadas bidimensionales (cruces en los rectángulos) de los cortes axial, sagital y coronal obtenidos a partir de las ARs en modo supervisado con *minsup* y *minconf* de 100%. $t=90\%$

evaluar la combinación de VAF con AE como método para la obtención de características con el objetivo de una extracción de ARs posteriormente para la clasificación de imágenes cerebrales. En este sentido, el rendimiento del clasificador basado en ARs, es evaluado como una herramienta para la detección temprana de AR, en términos de Acc, Sen y Spe. En el conjunto de experimentos consideramos un umbral creciente de número de ARs verificadas por el sujeto a clasificarse.

9.1.1 Resultados con SPECT

En todos los experimentos para atributos continuos, se seleccionan bloques activados tridimensionales de tamaño $9 \times 9 \times 9$ que incluyen vóxeles por encima de un umbral de intensidad (a_T) como se describió en la sección 6.2. La figura 9.1 muestra una representación del contorno del cerebro y la localización de las regiones activas con puntos, que se usan como entradas para el algoritmo de extracción de ARs. Al usar estos datos de preprocesado, automáticamente estamos reproduciendo la inspección visual de un médico pero de una forma más objetiva: regiones de interés comunes se buscan en controles y después se

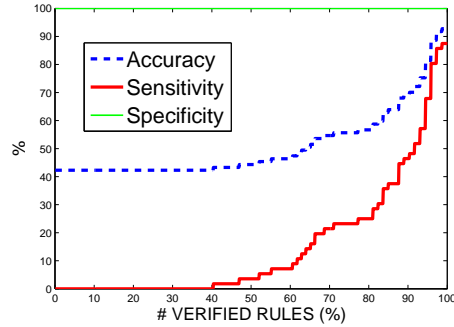


Figura 9.3: SPECT, método AR entre atributos continuos. Acc, Sen, Spe frente al porcentaje de reglas verificadas. Umbral de activación de bloque de 92%, con *minsup* y *minconf* de 100%.

aplican a los sujetos que se están testeando. Las coordenadas de los bloques se muestrean en términos de una malla tridimensional de tamaño $4 \times 4 \times 4$ para reducir la complejidad computacional del método. Las regiones activadas (figura 9.1) se obtienen y numeran para cada sujeto dentro de una imagen tridimensional calculada promediando todos los controles de la base de datos SPECT.

La figura 9.2 muestra los cortes transaxial, coronal y sagital de una imagen media SPECT que ilustra las localizaciones de la malla $4 \times 4 \times 4$ (cruces) y las regiones activadas involucradas en las ARs (cuadrados) en términos de antecedentes y consecuentes. Por lo tanto, los bloques tridimensionales se definen estableciendo un umbral de activación como una función del máximo nivel de intensidad de voxel (método AE) y están numerados ($I = i_1, i_2, \dots, i_m$) para describir transacciones $X \Rightarrow Y$ en la extracción de ARs. En particular, se probó experimentalmente que una región se considera de interés o *casi activada*, si más del 90% o una fracción de 650 voxeles del total de 729 están activados. De hecho, si se considera un umbral de activación de bloque más alto, los resultados de Acc, Sen y Spe son bastante buenos como se muestra en la figura 9.3 donde los bloques están al 92% de activación. Por otro lado, si se selecciona un umbral de bloque de una activación menor, los resultados de exactitud son más bajos con un incremento de tiempo computacional.

A continuación, analizamos el rendimiento de este método en términos de las medidas de significancia usadas en la extracción de ARs. En primer lugar, puede observarse que el número de reglas extraídas es mayor cuando los valores de *minsup* y *minconf* se reducen como se muestra en la figura 9.4, donde se obtiene un número mayor de ROIs (en comparación con el caso de la figura

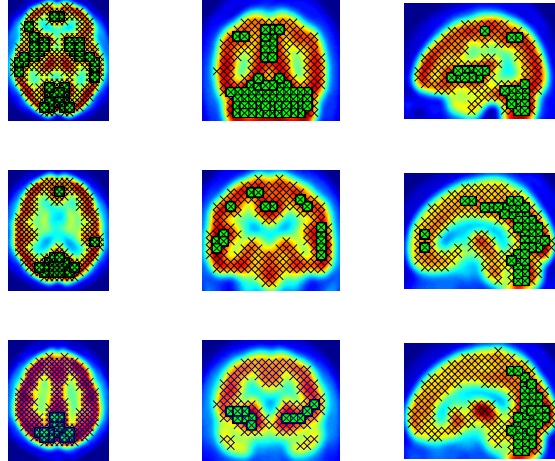
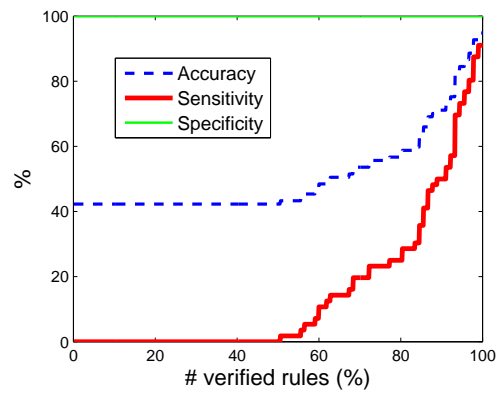


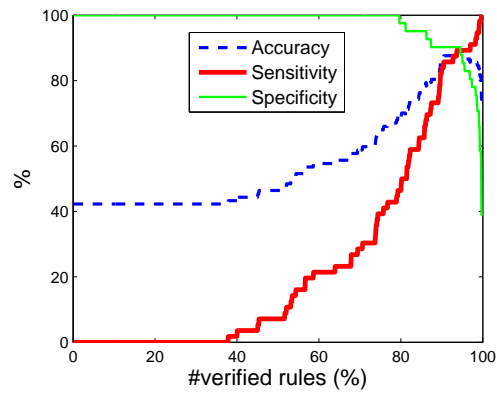
Figura 9.4: SPECT. Regiones activadas bidimensionales (rectángulos) de los ejes axial, sagital y coronal con $minsup$ y $minconf$ de 90%.

9.2 que representa los antecedentes y los consecuentes de las ARs cuando $minsup$ y $minconf$ son del 100%. Aunque el número de reglas extraídas se incrementa, destacamos que no todas son discriminativas ya que el soporte y la confianza de las mismas es bajo, es decir, muchas de estas reglas son redundantes ya que dependen de las propiedades específicas de las imágenes. Esto se muestra en las figuras 9.5 y 9.6 en la que $minconf$ y $minsup$ se reducen simultáneamente desde 100% a 80%. En este caso, la especificidad de la curva decrece a un porcentaje menor de reglas verificadas (n_0). Las ARs extraídas son discriminativas para el diagnóstico precoz de la AD cuando se seleccionan los $minsup$ y $minconf$ más altos, es decir, 100%, en caso contrario incluimos cientos de reglas dependientes de los datos que no modelan de forma efectiva el patrón normal. Además se corrobora esta suposición con el decremento de la especificidad para un valor de reglas verificadas n_0 en las figuras (9.5(b) con $minsup$, $minconf$ de 95% a $n_0=95\%$ de reglas verificadas, 9.6(a) $minsup$ and $minconf$ de 90% a $n_0=88\%$ y en 9.6(b) con $minsup$, $minconf$ de 80% a $n_0=79.5\%$).

Como conclusión, si los valores de $minsup$ y $minconf$ son menores que el 100%, este método provee una mayor habilidad para detectar un valor específico n_0 ocurriendo que cuando se considera un valor $n > n_0$, la especifi-

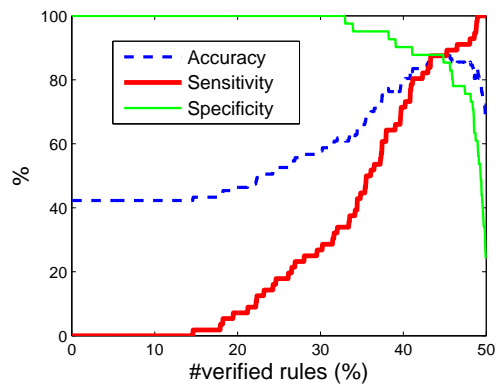


(a)

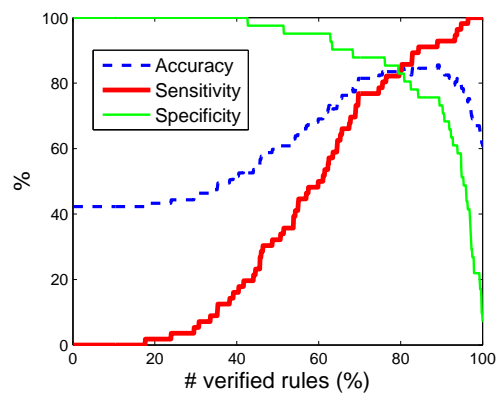


(b)

Figura 9.5: Acc, Sen, Spe frente al % ARs verificadas (SPECT), $t=90\%$ para $minsup$, $minconf$: a) 100%, b) 95%



(a)



(b)

Figura 9.6: Acc, Sen, Spe frente al % ARs verificadas (SPECT), $t=90\%$ para $minsup$, $minconf$: a) 90% b) 80% .

cidad decrece notablemente. Además, esto puede conllevar un incremento en el coste computacional de un sistema CAD con una penalización en la tasa de exactitud máxima (89.69% en la figura 9.5(b), 87.63% en la figura 9.6(a) y 85.57% en la figura 9.6(b)) ya que las reglas no discriminativas también están incluidas en el proceso de clasificación. Nótese que en general el número de reglas extraídas y las exactitudes máximas obtenidas para *minsup* y *minconf* de 100%, 95%, 90%, 80% son de: a) 29756 reglas, 94.87%, b) 84396 reglas, 89.69%, c) 146768 reglas, 87.63% and d) 288190 reglas, 85.57%. A la vista de las figuras 9.5 y 9.6, se destaca que extrayendo reglas al más alto porcentaje de *minsup* y *minconf*, se logran los mejores resultados en términos de exactitud, especificidad y sensibilidad (94.87%, 100% and 91.07% respectivamente). En este caso el coste computacional se reduce también, debido a que el análisis se enfoca en las regiones cerebrales más relevantes para el diagnóstico de la AD (véase figura 9.2).

Análisis de curva ROC para SPECT

La curva ROC tiene como objetivo comparar el compromiso entre la *Sen* (Fracción de verdaderos positivos) y $1 - Spe$ (Fracción de falsos positivos) para describir de una forma completa la detectabilidad de la enfermedad [Metz, 1978] como se explica en la sección 5.6.1. La figura 9.7 muestra las curvas ROC para diferentes valores de *minsup* y *minconf* (de 100% a 80%) que evalúan el compromiso entre la sensibilidad y la especificidad conforme se incrementa el umbral de ARs verificadas por cada sujeto. Se ha establecido una comparación con otros métodos CAD recientemente publicados tales como VAF [Stoekel et al., 2001] y PCA [López et al., 2009a] (de los que se aportan la curva ROC y el punto de operación) y GMM [Górriz et al., 2009] combinados con SVM, que se representan por su punto de operación. La curva ROC del método propuesto se desplaza arriba y a la izquierda mejorando el método de partida utilizado en este capítulo como se muestra en la figura 9.7.

Cuando *minsup* y *minconf* se seleccionan del 100%, todos los controles se clasifican adecuadamente, es decir la *Spe* es de 1 mientras que la *Sen* alcanza 0.91 aproximándose al sistema ideal en la curva ROC ($Sen, Spe = 1$) en el que cada sujeto se clasificaría correctamente. En esta suposición, nuestro sistema CAD basado en AR clasifica correctamente todos los controles (41) y 51 (de 56) pacientes AD. De los 5 pacientes AD mal clasificados, 1 (de 22) es AD2 y 4 (de un total de 30) son AD1, es decir aunque 26 AD1 se clasifican bien, el patrón AD1 es aún un reto para el diagnóstico precoz de la AD. Si sólo consideramos el caso NOR frente a AD1, los 41 controles y 26 AD1 (de 30) se clasifican bien siendo los ratios de precisión del método $Acc = 94.37%$, $Sen = 86.67%$, $Spe = 100%$ que aún suponen un reto para el

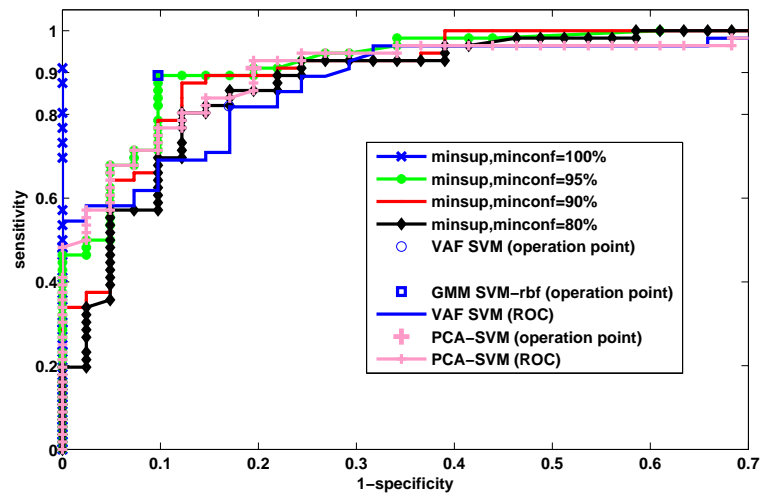


Figura 9.7: Base de datos SPECT. Análisis ROC: método AR para atributos continuos con diferentes porcentajes seleccionados de $minsup$ (100%, 95%, 90%, 80%) y $minconf$ (100%, 95%, 90%, 80%) en la extracción de reglas. Comparación con otros métodos recientemente presentados que se representan por sus puntos de operación. Obsérvese el valor de pico de $Acc = 94.87\%$, con $Sen = 91.07\%$ y $Spe = 100\%$ (marcas en cruz), para $minsup = minconf = 100$. La AUC obtenida para cada ROC es de: método AR ($minsup$: AUC) 100%: 0.9554, 95%: 0.9371, 90%: 0.9233, 80%: 0.9027; VAF SVM: 0.8993; PCA SVM: 0.9177

<i>minconf</i> , <i>minsup</i>	100%	95%	80%
% ARs verificadas	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)
70	53.71 / 19.64 / 100	58.76 / 28.57 / 100	81.44 / 76.79 / 87.80
90	71.13 / 50 / 100	86.60 / 83.93 / 100	84.54 / 92.86 / 73.17
100	94.85 / 91.07 / 100	74.23 / 100 / 39.02	60.82 / 100 / 0.07
	VAF	PCA	GMM
	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)
	83.51 / 83.93 / 82.93	86.56 / 91.07 / 80.49	89.69 / 90.24 / 89.29

Tabla 9.1: Medidas estadísticas del rendimiento de los métodos de selección de características basados en reglas usando varias configuraciones (*minconf*, *minsup* y % reglas verificadas). VAF, GMM y PCA se utilizan como referencia. Los parámetros VAF: clasificador SVM lineal, parámetros GMM: $\sigma=6$ clasificador RBF-SVM con 8 componentes y parámetros PCA: $\sigma=6$ y clasificador RBF-SVM con 16 componentes.

diagnóstico y que siguen representando un gran avance en el campo.

Si un sujeto verifica un número mayor o igual de ARs que un umbral de clase, se clasifica como control y en otro caso como AD. Esto significa que conforme crece el umbral de reglas verificadas, un número menor de casos se clasificaría como control, es decir la Spe debería decrecer. Esto no ocurre sin embargo cuando se selecciona el más alto *minsup* y *minconf* (100%) en la extracción de ARs, es decir, cuando las reglas modelan perfectamente el patrón normal ya que son las reglas más discriminativas para la enfermedad de Alzheimer. Los mejores resultados de clasificación se obtienen cuando se verifica el 100% de las reglas. Estos resultados están en concordancia con el comportamiento esperado del sistema ya que se asume que los sujetos normales tienen un patrón común de SPECT y verifican la mayoría de las reglas extraídas. Este hecho apoya la idea de que es preferible usar sólo la extracción de reglas en un conjunto de controles. En la tabla 9.1 resumimos las medidas estadísticas obtenidas con los métodos propuestos y las comparamos con sistemas CAD recientemente publicados en [Segovia et al., 2010; Stoeckel et al., 2004]. Se concluye que el análisis de las curvas ROC muestra que el sistema CAD propuesto basado en la extracción de las reglas de asociación alcanza el mejor compromiso entre la sensibilidad y la especificidad desplazando el punto de operación arriba y a la izquierda en el espacio ROC [Metz, 1978]. Como se espera, los mejores resultados se obtienen con la técnica propuesta de ARs

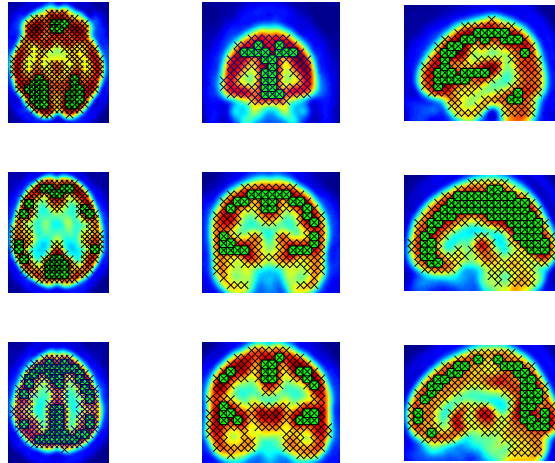


Figura 9.8: Base de datos PET Cartuja. ROIs activadas bidimensionales (cruces en los rectángulos) de los cortes axial, sagital y coronal obtenidos a partir de las ARs en modo supervisado con *minsup* y *minconf* de 100%).

para los más altos *minsup* y *minconf* (de 100%).

9.1.2 Resultados para PET Cartuja

Para el análisis de esta base de datos, se han utilizado ROIs del mismo tamaño v que para SPECT y con la misma activación t , para mostrar la robustez de la metodología propuesta. En la figura 9.8 se representan las ARs en los cortes axial, sagital y coronal cuando *minsup* y *minconf* son del 100%. El número de reglas representadas es superior respecto a SPECT y está repartido prácticamente por todo el cerebro dado que los pacientes AD de esta base de datos se encontraban en un estado bastante avanzado de la enfermedad. En este sentido, el análisis precoz mediante imágenes funcionales se facilita frente a otras bases de datos. Sin embargo, con el método de ARs en todas las bases de datos los resultados fueron siempre superiores al 90%, lo que apoya la robustez de esta técnica para el análisis de imagen funcional.

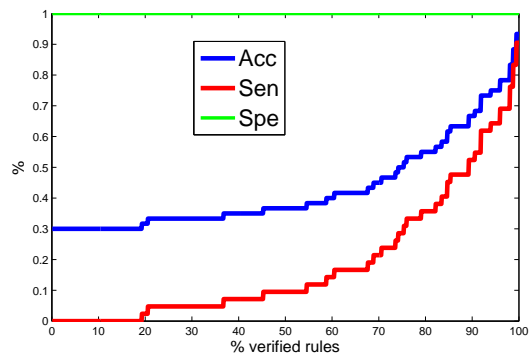
La figura 9.9(a) muestra la exactitud, sensibilidad y especificidad frente al porcentaje del número de reglas verificadas. Dado que *minsup* y *minconf* es del 100%, los máximos ratios (Acc=93.33%, Spe=100%, Sen=90.48%) se alcanzan en el 100% de reglas verificadas (es decir, 82082 reglas). En la figura 9.9(b)

se muestra la curva ROC del método propuesto en comparación con PCA y VAF combinados con SVM. El método basado en ARs mejora sensiblemente los otros ya que el punto de operación se desplaza más arriba y a la izquierda. Estos resultados concuerdan con la AUC que aparece descrita en el pie de la figura 9.9.

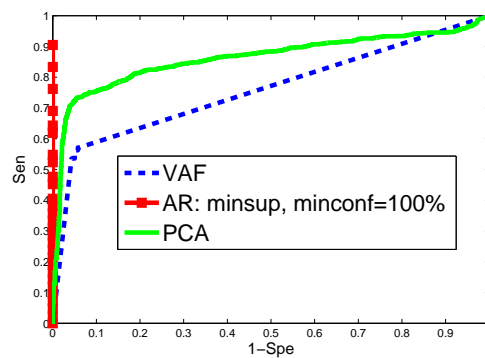
9.1.3 Resultados para PIB y FDG

Tradicionalmente, los cambios en el metabolismo de la glucosa en el cerebro humano se han medido mediante la imagen FDG-PET [Kubota et al., 2011] que ha sido utilizada para visualizar el hipometabolismo en pacientes con AD. Como se sugiere en [Shin et al., 2010], el uso combinado de radiotrazadores en los mismos sujetos puede proporcionar información valiosa para estudiar los patrones patológicos relacionados con las placas de amiloide, neurofibrillas e hipometabolismo de la glucosa en AD. El análisis basado en vóxel de PIB-PET [Illán et al., 2010] en pacientes con AD comparados con sujetos normales mostró que los primeros presentan una retención significativa de PIB en las regiones cerebrales conocidas por tener un gran depósito de placas de amiloide (por ejemplo en los cortes frontal, parietal, temporal y posterior cingulado/precuneus). El análisis basado en vóxel de FDG-PET en los mismos sujetos reveló que las regiones cerebrales con hipometabolismo de glucosa en pacientes con AD se solapan con regiones de alta concentración PIB [Shin et al., 2010] como se muestra en la figura 9.10. Se concluye que las imágenes PET utilizando estos radiotrazadores combinados en los mismos sujetos pueden contribuir al desarrollo y prueba de los medicamentos que tienen como objetivo mitigar la patología del amiloide, la patología tau *y/o* el metabolismo de energía. En este trabajo se combina la información de los biomarcadores PIB y FDG para obtener información complementaria al analizar imágenes funcionales mediante ARs.

Esta sección pretende estudiar los beneficios de utilizar varios biomarcadores conjuntamente (o separadamente) en la evaluación de una base de datos utilizando las reglas de asociación entre atributos continuos. Así mismo, se comprueba qué biomarcador funciona mejor [Chaves et al., 2012a] con cada grupo de evaluación: CTRL frente a AD o CTRL frente a MCI. El método consiste en las mismas 4 etapas explicadas en la sección 6.2: enmascarado, estimación de la activación para la selección de regiones de interés (ROIs), extracción de ARs y clasificación. Después del preprocesado, un promedio de controles se aplica para descartar regiones ruidosas. Cada imagen PET se divide en vóxeles cúbicos tridimensionales de tamaño $9 \times 9 \times 9$ con solapamiento $4 \times 4 \times 4$ que están dentro de la máscara. En la fase de estimación de la acti-



(a)



(b)

Figura 9.9: Base de datos PET Cartuja. a) Acc, Spe y Sen en función del porcentaje de reglas verificadas (total de reglas=82082). b) Curva ROC en comparación con otros métodos publicados recientemente. AUC: VAF(0.7592), PCA(0.8683), AR(0.9158).

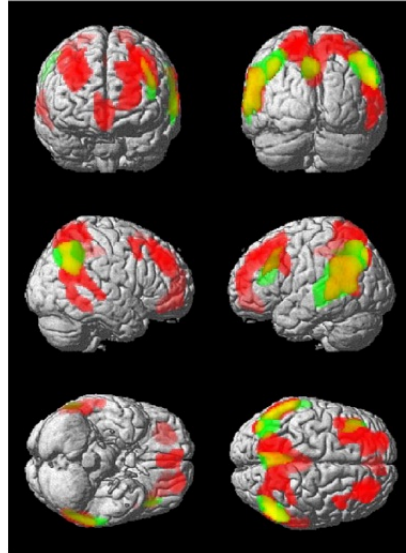


Figura 9.10: Comparación de las imágenes PIB-PET (rojo) con FDG-PET (verde). En amarillo se representa el área donde se solapan los mapas t de SPM [Shin et al., 2010].

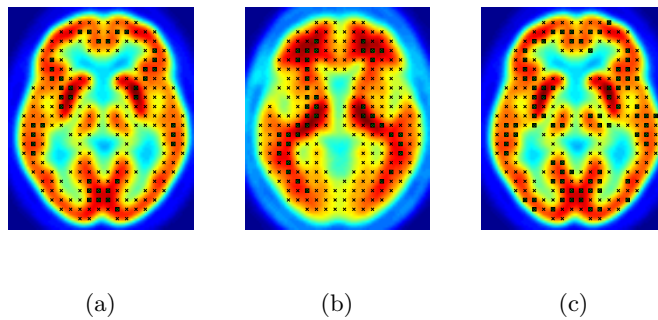


Figura 9.11: Base de datos FDG-PIB. ARs pintadas como cuadrados en los cortes coronales para un $minsup, minconf=100\%$ para la base de datos PET: a) FDG, b) PIB, c) FDG más PIB (sobre una imagen media FDG de los controles). Grupo AD-NOR

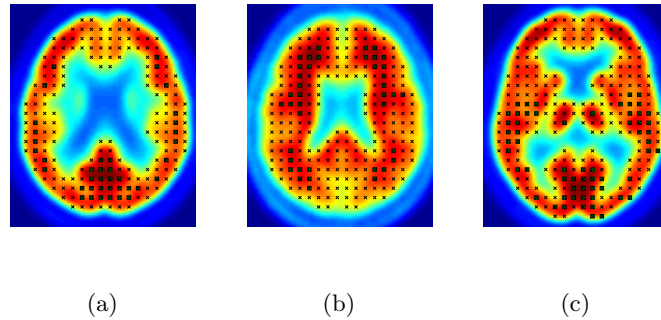


Figura 9.12: Base de datos FDG-PIB, grupo MCI-NOR. ARs pintadas como cuadrados en los cortes coronales para un $minsup, minconf=100\%$ para la base de datos PET: a) FDG, b) PIB, c) FDG más PIB (sobre una imagen media FDG de los controles).

vación, un vóxel está activado definiendo ROIs si el ratio de vóxeles activados dentro de él es superior a un umbral del 90%. Los experimentos consideraron el entrenamiento y test del sistema en dos conjuntos diferentes:

1. CTRL frente a AD.
2. CTRL frente a MCI.

El sistema CAD propuesto se analiza usando separadamente la base de datos con el biomarcador $[^{18}F]$ -FDG o PIB y la combinación de ambos. Para el análisis de la integración de biomarcadores, las ROIs obtenidas tras la fase de estimación de activación con FDG o PIB se combinan sin repetición. Esta combinación de ROIs obtenidas se utiliza como entrada en la fase de extracción de reglas, que se realiza en términos del algoritmo *A priori* tomando como entrada los valores de $minconf$ y $minsup$. En la fase de test, se comprueba el número de reglas verificadas por cada sujeto. Si un sujeto verifica igual o mayor número de reglas que un umbral de clase, se clasifica como CTRL y en otro caso como AD (para el conjunto 1) o MCI (para el conjunto 2).

El sistema propuesto fue analizado en términos de la Acc, Sen y Spe. En primer lugar se definen ROIs de tamaño $9 \times 9 \times 9$ por estimación de activación. Las figuras 9.11 y 9.12 muestran para el conjunto primero y segundo respectivamente, los cortes coronales de la máscara FDG (a y c) y de la máscara PIB (b) ilustrando las posiciones $4 \times 4 \times 4$ (cruces) y las regiones activadas involucradas en las ARs (cuadrados) en términos de los antecedentes y consecuentes de las reglas. Se observa que cuando se combinan los biomarcadores FDG y

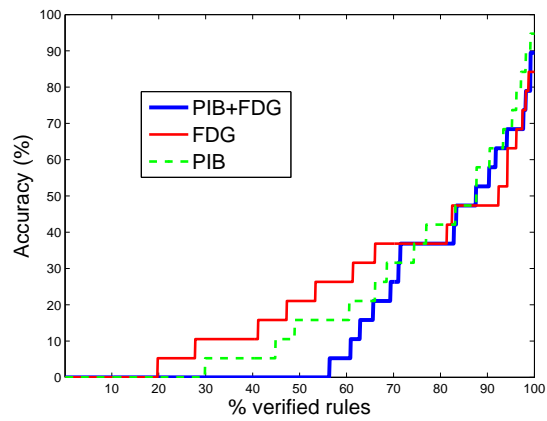
PIB en la máscara FDG (figuras 9.11(c) y 9.12(c)), aparecen dibujadas más cruces ya que se extraen nuevas reglas mejorando resultados obtenidos sólo con ROIs FDG.

Las figuras 9.13(a) y 9.13(b) muestran los resultados de exactitud (para el conjunto 1 y 2 respectivamente) frente al porcentaje de reglas verificadas cuando se utiliza por separado el biomarcador FDG, PIB o ambos combinados. Dado que el *minsup* y *minconf* utilizados son del 100% y que las reglas obtenidas son las más discriminativas para el diagnóstico de AD, las exactitudes máximas se obtienen al alcanzar el 100% de las reglas verificadas. Para el conjunto 1, los resultados máximos obtenidos son (Acc/Sen/Spe) de (92.11%/84.21%/100%) para FDG y (97.37%/94.74%/100%) para PIB. La combinación de biomarcadores mejora los resultados obtenidos con FDG alcanzando (94.74%/89.47%/100%). Para el conjunto 2, los resultados fueron mejores en PIB (90.48%/87.69%/100%) respecto a FDG (77.38%/70.7%/100%) mostrando así la misma tendencia que para el conjunto 1. Sin embargo, la combinación de biomarcadores *FDG + PIB* es muy positiva en la evaluación de este conjunto CTRL frente a MCI dado que se obtienen resultados máximos de (92.85%/90.77%/100%) mejorando los obtenidos con sólo un biomarcador. En comparación con otros métodos publicados, la técnica propuesta basada en reglas mejora PCA en combinación con SVM, en las que se alcanzaron exactitudes máximas de 86.11% [Illán et al., 2010].

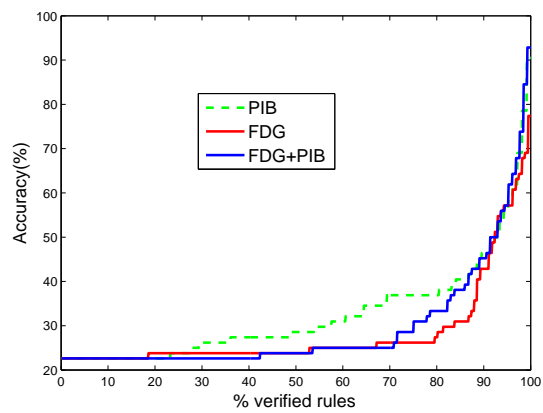
Los resultados obtenidos están en concordancia con las curvas ROC, que se muestran en las figuras 9.14 para el conjunto 1 (el tramo superior aparece ampliado para mayor claridad en la figura 9.14.b) y en la figura 9.15 (el tramo superior aparece ampliado para mayor claridad en la figura 9.15.b) para el conjunto 2 usando 100% de *minsup* y *minconf* concluyendo que el sistema CAD propuesto basado en ARs alcanza el mejor compromiso entre la sensibilidad y la especificidad desplazando el punto de operación arriba y a la izquierda en el espacio ROC [Metz, 1978]. Para el conjunto 1, se produjeron los mejores resultados usando el método propuesto con el biomarcador PIB; mientras que para el conjunto 2, la combinación del biomarcador PIB con FDG favoreció la mejora de los mismos.

9.2 ARs entre atributos discriminantes

Como se explicó en el capítulo 6, la introducción de ARs entre atributos discriminantes favorece la mejora del coste computacional, dado que el número de coordenadas entre las que se establecen asociaciones se reduce al aplicar un umbral basado en FDR. Se realizaron varios experimentos para testear la

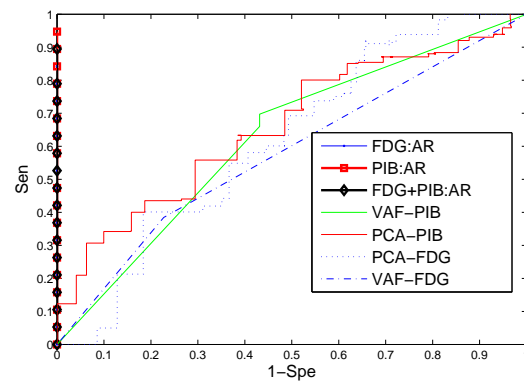


(a)

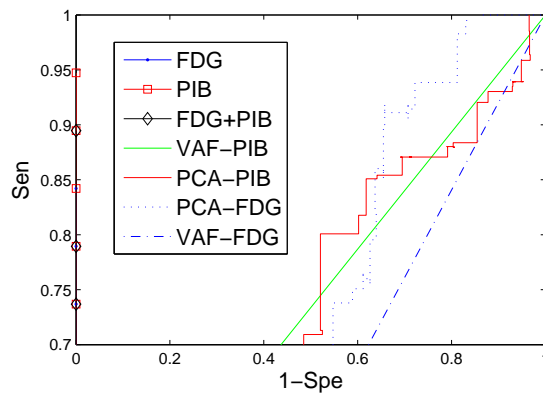


(b)

Figura 9.13: Base de datos FDG-PIB Exactitud (%) frente al porcentaje de reglas verificadas al utilizar FDG, PIB o combinación. a) CTRL-AD, b) MCI-NOR.

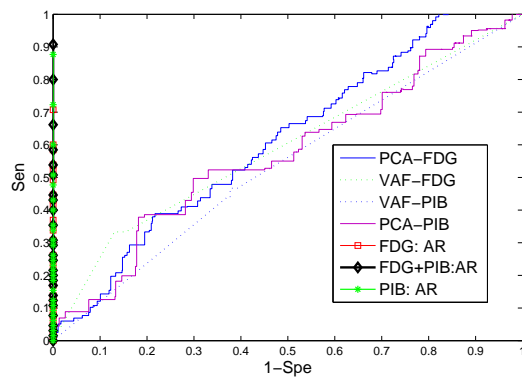


(a)

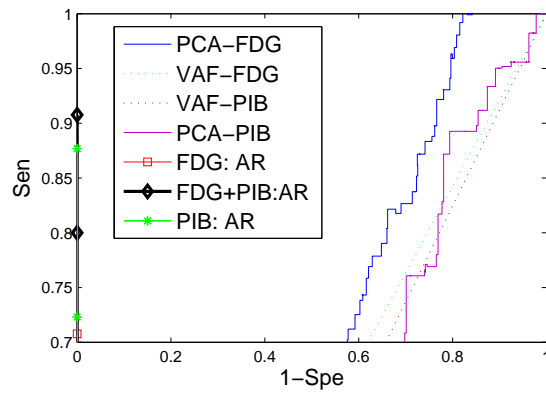


(b)

Figura 9.14: Base de datos FDG-PIB, método ARs entre atributos continuos. ROC del método propuesto en comparación con otros métodos publicados recientemente. a) NOR frente a AD b) ampliación del tramo superior.



(a)



(b)

Figura 9.15: Base de datos FDG-PIB, método ARs entre atributos continuos. ROC del método propuesto en comparación con otros métodos publicados recientemente. a) NOR frente a MCI b) ampliación del tramo superior.

	$h = 0.2$	$h = 0.4$	$h = 0.6$	$h = 0.8$
PET: Max Num reglas extraídas	7482	5256	1332	210
PET: Acc (%)	91.33	90.67	88.67	84.00
SPECT: Max Num reglas extraídas	67860	132	0	0
SPECT: Acc (%)	92.68	65.98	-	-

Tabla 9.2: Comparación de la máxima exactitud y número de ARs extraídas y obtenidas para las bases de datos SPECT y PET para diferentes umbrales de la máscara de selección basada en vóxeles FDR.

fiabilidad del clasificador propuesto. Usamos dos modalidades de imagen de bases de datos (SPECT y FDG-PET de ADNI) en el análisis para resaltar la capacidad de generalización del método.

Las ROIs se seleccionaron como bloques tridimensionales de tamaño $5 \times 5 \times 5$ localizadas dentro de la correspondiente máscara FDR SPECT o PET. Las coordenadas centrales de los bloques tridimensionales se restringen a una malla tridimensional de $2 \times 2 \times 2$. El hecho de presentar bloques reducidos en tamaño frente a la sección anterior entre atributos continuos, se debe a que en primer lugar queremos generalizar el método a ambas bases de datos y ADNI-PET debido a que muestra una activación reducida frente a SPECT en las ROIs trabaja mejor con bloques de tamaño reducido. Aunque se establecieron pruebas para diferentes umbrales, se seleccionó como el más adecuado un umbral FDR de $h = 0.2$.

En la fase de extracción de características, un bloque tridimensional se define como activado si más de una fracción (50 vóxeles) del total (125 vóxeles) están activados. Este umbral representa un 40% de vóxeles activados y permite sólo que los volúmenes cerebrales con una alta activación se consideren ROIs. Concluimos que para ambas bases de datos, el umbral de activación se reduce cuando se aplica la máscara FDR ya que sitúa en sí misma los vóxeles más discriminativos para la detección temprana de la AD. La figura 9.16 muestra los cortes coronal, transaxial y sagital de una imagen media de los controles PET ilustrando los vóxeles dentro de la máscara FDR (cruces) así como las ROIs activadas representando las ARs (cuadrados), que se amplían en la parte inferior de la figura 9.16 para mayor claridad. Estas ROIs se corresponden con aquellas regiones que están comúnmente afectadas por la AD, es decir el giro cingulado posterior y precuneiforme, así como la región temporoparietal.

Las ARs se extraen de los controles usando el *minsup* y *minconf* más alto, es decir 100%, para extraer el patrón más común en los sujetos normales y proporcionar una habilidad de detección mejorada en el diagnóstico precoz de

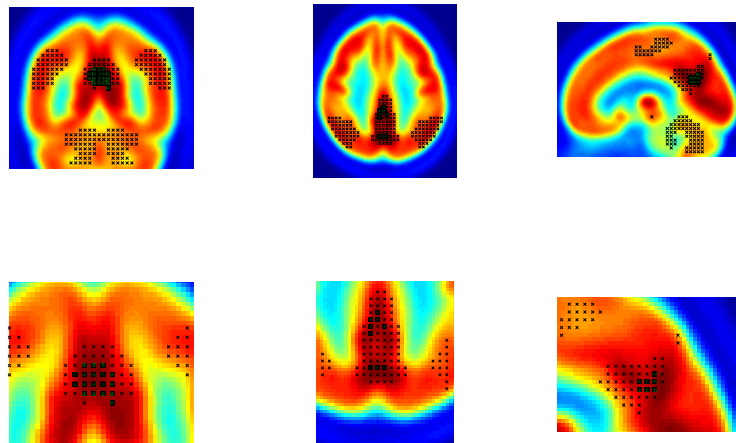


Figura 9.16: Regiones activadas bidimensionales de los cortes coronales, axiales y sagitales para ARs entre atributos discriminantes. En la segunda fila se realiza una ampliación de los cortes cerebrales de la primera fila para representar las regiones cerebrales discriminativas de las que se extraen las ARs.

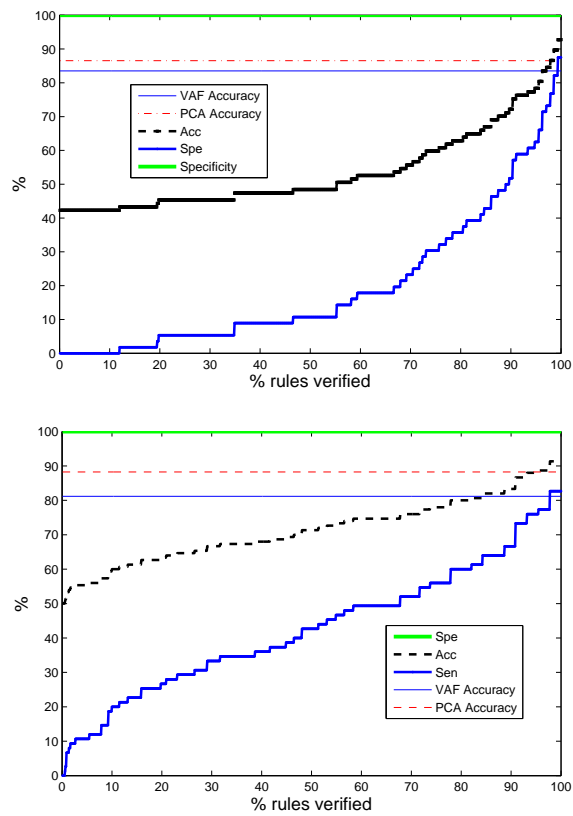


Figura 9.17: Valores de Acc, Sen, Spe (método ARs entre atributos discriminantes) como una función del umbral de reglas verificadas en (%) para a)SPECT b)PET. Comparación con otros métodos publicados recientemente.

la AD. Los experimentos consideraron un umbral creciente para el número de reglas verificadas por cada sujeto como se muestra en la figura 9.17 a) SPECT, b) PET. Se muestra la exactitud, sensibilidad y especificidad del sistema CAD propuesto como una función del umbral de clasificación (número de reglas verificadas por cada sujeto). Además, se muestra una comparación con otros métodos recientemente publicados incluyendo VAF [Fung and Stoeckel, 2007] o PCA en combinación con un clasificador SVM. Los valores de Acc, Spe y Sen convergen a un máximo de 92.78%, 100% y 87.5% para SPECT y 91.33%, 100% y 82.67% para PET, respectivamente. La exactitud de clasificación mejora cuando se incrementa el umbral de reglas verificadas para cada sujeto ya que se debe satisfacer un número mayor de reglas para clasificar cualquier sujeto que tenga un patrón más parecido a un normal, por lo que más pacientes AD se detectan correctamente. En particular, 67860 reglas representan el número máximo de reglas verificadas (denotadas por 100% en el eje x de la figura 9.17) para SPECT, mientras que para PET, este número alcanza 7482 reglas (véase tabla 9.2). Se concluye que el tamaño del área derivada del análisis FDR en SPECT es mayor que en PET donde los déficits funcionales están más localizados en regiones cerebrales bien conocidas. Por lo tanto, el número de ROIs usadas como entradas para el proceso de extracción de ARs es mayor para SPECT debido al tamaño del mapa de activación de la máscara. Se observa también que los valores de especificidad para este sistema CAD alcanzan 100%, es decir, todos los controles se clasifican correctamente cuando los valores de *minsup* y *minconf* en el proceso de extracción de reglas son iguales al 100%. Estos resultados están en concordancia con el comportamiento esperado del sistema ya que se asume que los sujetos normales tienen un patrón de imagen común y que verifican la mayoría de las ARs extraídas. Además hemos incluido el estudio del compromiso entre el coste computacional y la habilidad de detección para analizar las demandas computacionales. La tabla 9.2 muestra el coste computacional en términos del número de reglas extraídas para un umbral FDR determinado h , así como la exactitud máxima alcanzada. Este umbral se utiliza como un criterio de selección de vóxel y permite reducir la dimensión de las variables de entrada en el proceso de extracción de ARs. La exactitud del sistema mejora reduciendo el umbral h a costa de incrementar el número de ARs extraídas y como consecuencia, la complejidad computacional. El mejor compromiso entre los resultados de clasificación y el coste computacional se obtiene para un umbral $h = 0.2$.

Se puede concluir que el sistema propuesto mejora sistemas CAD recientemente desarrollados basados en VAF y PCA [Álvarez et al., 2009], [Illán et al., 2011b], [López et al., 2009a] que alcanzaron exactitudes de clasificación

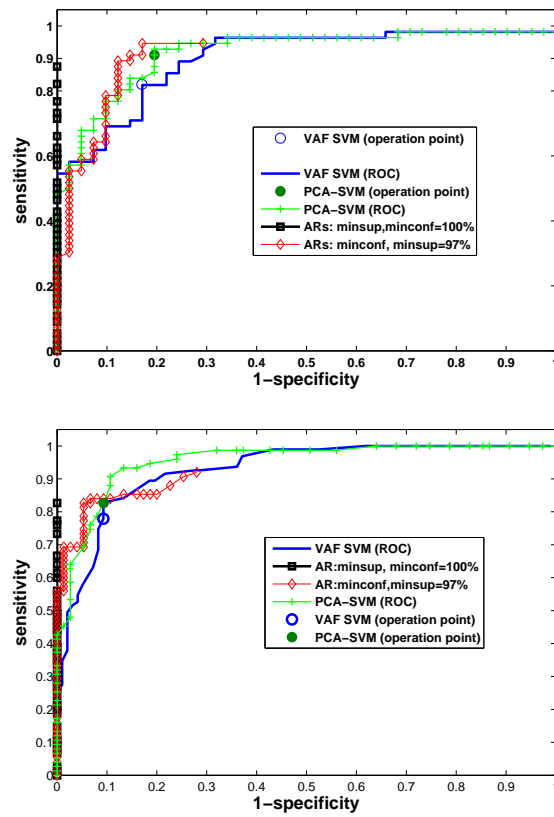


Figura 9.18: Curva ROC (método ARs entre atributos discriminantes) para a) SPECT y b) PET. Comparación con otros métodos recientemente publicados.

Intervalos de discretización	1	2	3	4	5
1	–	44.33	64.95	96.91	96.91
2	44.33	44.33	64.95	51.55	96.91
3	64.95	64.95	79.38	74.23	67.01
4	96.91	51.55	74.23	96.91	95.88
5	96.91	96.91	67.01	95.88	96.91

Tabla 9.3: Exactitud (%) tomando la quinta subimagen como máscara del total ($K_1 = 5$) y combinando diferentes intervalos en la fase de discretización ($K_2 = 5$). SPECT, ARs entre atributos discretos.

de imágenes de 83% o 86% para SPECT y 81.18% o 88.24% para PET, respectivamente. Las curvas ROC han demostrado ser muy efectivas para la evaluación de sistemas CAD como se explica en la sección 5.6.1. La figura 9.18 muestra las curvas ROC del sistema CAD (a) para SPECT y b) PET) basado en FDR-AR así como los métodos para comparación VAF y PCA. El método propuesto al 100% de *minconf* y *minsup* mejora algunos métodos recientemente publicados como VAF o PCA dado que su ROC está desplazada a la izquierda y hacia arriba en el espacio ROC. De hecho, el método propuesto alcanza 87.5% y 82.67% de sensibilidad para las modalidades SPECT y PET, respectivamente, y el 100% de especificidad. La figura 9.18 también muestra la curva ROC del método propuesto para el 97% de *minconf* y *minsup*. En este caso, el método basado en AR aporta un menor ratio de exactitud que la configuración previa del modelo, sin embargo este análisis proporciona una información útil para la selección de parámetros dado que cuanto mayor es *minsup* y *minconf*, mejores resultados se obtienen. Una de las ventajas del método propuesto es que se reduce la complejidad de la decisión que consiste en un proceso de comprobación del número de reglas verificadas que se están estudiando. Esto se debe al hecho de restringir que las ARs estén definidas sobre volúmenes cerebrales dentro de una máscara FDR, reduciendo de este modo la complejidad del problema. En este caso, sólo se incluyen las áreas cerebrales relevantes en la extracción de reglas evitando la evaluación del algoritmo *Apriori* en regiones activadas irrelevantes, como el cerebelo.

9.3 ARs entre atributos discretos

Como se justificó en el capítulo 5 y 6, el hecho de utilizar atributos discretos frente a continuos umbralizados, favorece la automatización del problema de las ARs. La discretización de atributos ha de realizarse previa a la selección

de regiones, típicamente dividiendo el rango o subrango en subintervalos. En esta sección, llevamos a cabo algunos experimentos para probar la fiabilidad del sistema CAD propuesto cuyo rendimiento se evalúa en términos de su Acc, Sen y Spe. Los bloques tridimensionales se escogen de tamaño $9 \times 9 \times 9$ (como en el caso de atributos continuos) centrados en las coordenadas de la máscara seleccionada y se lleva a cabo un proceso de discretización para la selección de ROIs. Los centros de las coordenadas de los bloques 3D se restringen a pertenecer al grid tridimensional de $4 \times 4 \times 4$.

El enmascarado se lleva a cabo como se ha explicado en la sección 6.4, mediante un histograma de imagen y eligiendo la máscara que mejor actúe sobre cada base de datos (SPECT o PET ADNI). En segundo lugar, la activación de las ROIs elegidas como entrada para *Apriori*, se seleccionan mediante la combinación de intervalos de discretización (la activación de las regiones en lugar de ser elegida mediante un umbral t , es seleccionada automáticamente mediante la combinación de intervalos discretizados). Tanto el enmascarado como la discretización se llevan a cabo con $K_1, K_2 = 5$ para aportar robustez al método en ambas bases de datos. En lo que al enmascarado se refiere, el histograma muestra una distribución de igual anchura cubriendo K_1 rangos. Si pocos rangos (o muchos por el contrario) se utilizan, el histograma puede desorientar. $K_1 = 5$ seleccionado tiene la pretensión de mostrar este compromiso a la vista de los resultados con ambas bases de datos.

Referente a la discretización (mediante intervalos de igual tamaño), se seleccionó $K_2 = 5$ como un compromiso entre el coste computacional y los resultados de exactitud: conforme K_2 se incrementa, el número de ROIs seleccionadas por discretización decrece, acelerando de esta forma el proceso de extracción de ARs pero descartando regiones importantes para el diagnóstico precoz de la AD. Por otro lado, conforme K_2 decrece, el coste computacional se reduce y no mejora los ratios de clasificación en este caso. Se destaca que la calidad de las ARs para la clasificación es mejor cuando *minsup* y *minconf* se seleccionan del 100% conforme el patrón de controles se distingue mejor del de AD. Cuando se seleccionan intervalos discretizados (normalmente los más bajos) no tienen poder predictivo suficiente y tampoco se extraen ARs (véase este efecto en las tablas 9.3 y 9.4 para los intervalos discretizados 1 y 2).

9.3.1 Experimentos con la base de datos SPECT

En el proceso de enmascarado, se selecciona la cuarta subimagen (de $K_1 = 5$) (véase figura 6.8) ya que proporciona los mejores resultados de clasificación. La tabla 9.3 muestra las combinaciones de intervalos cuando $K_2 = 5$ y se demuestra que cuando los intervalos discretizados mayores se incluyen en la

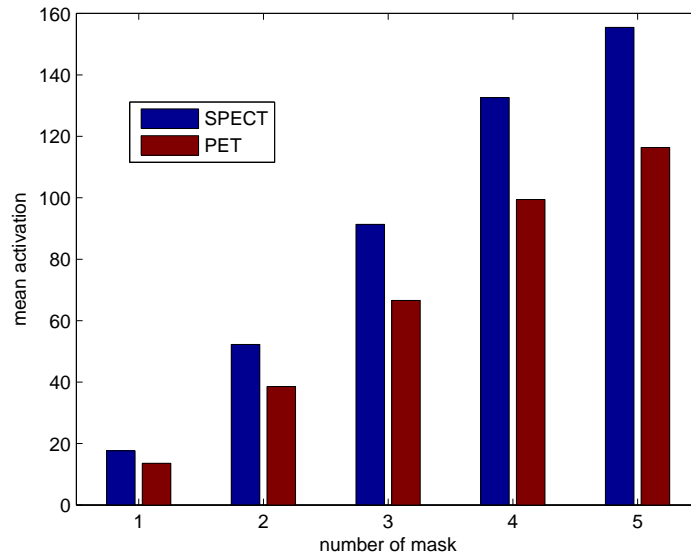


Figura 9.19: Promedio de características con los valores de la activación media para las bases de datos SPECT y PET.

combinación, los resultados de exactitud mejoran con grandes probabilidades. De hecho, cuando se combinan el 5^o con el 1^o, 2^o o 5^o se obtienen los mejores resultados de exactitud 96.91% (94.64% sensibilidad y 100% especificidad), mejorando la técnica basada en ARs continua [Chaves et al., 2011a] con resultados ($Acc = 94.87\%$, $Sen = 91.07\%$, $Spe = 100\%$).

9.3.2 Experimentos con la base de datos PET

Hablando de forma general, PET está menos activado que SPECT. Esto se demuestra en la figura 9.19 que muestra el promedio de las características de intensidad media para las bases de datos PET y SPECT. Se resalta que la 4^a máscara en SPECT y la 5^a en PET (véase figura 6.8) tienen el valor de intensidad media más similar (115.38 para SPECT y 132.59 para PET) lo que garantiza una mejor habilidad de detección. La tabla 9.4 muestra las combinaciones discretizadas de intervalos para PET cuando la máscara presentada en la figura 6.8(1) se aplica, alcanzando tasas de ($Acc = 92\%$, $Sen = 84\%$ and $Spe = 100\%$) mejorando el caso continuo basado en ARs ($Acc = 91.33\%$, $Sen = 82.6\%$, $Spe = 100\%$) [Chaves et al., 2012c]. De la misma forma que para la base de datos SPECT, se obtuvieron los mejores

Intervalos de discretización	1	2	3	4	5
1	–	–	86.67	91.33	78
2	–	–	86.67	57.33	78
3	86.67	86.67	88	64	59.33
4	91.33	57.33	64	91.33	92
5	78	78	59.33	92	78

Tabla 9.4: Tasa de exactitud (%) tomando la 5^a subimagen ($K_1 = 5$) como máscara y combinando diferentes intervalos de la fase de discretización ($K_2 = 5$). FDG-PET, ARs entre atributos discretos.

resultados para los intervalos discretizados más altos ya que muestran más activación y mejor habilidad de detección para la AD.

9.3.3 Estudio ROC

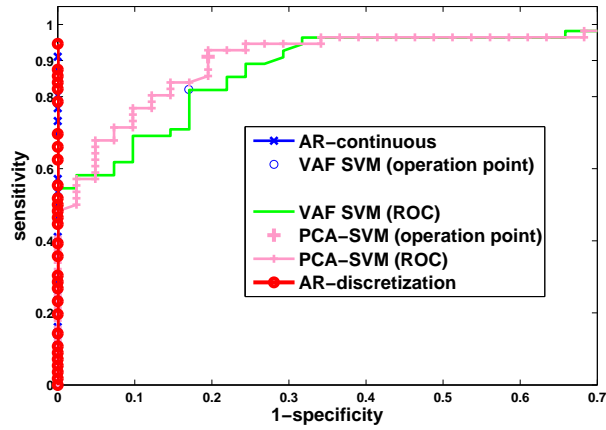
La figura 9.20 muestra las curvas ROC del sistema CAD propuesto basado en AR y discretización para SPECT (figura 9.20(a)) y PET (figura 9.20(b)). La curva ROC para el método propuesto se dibuja para SPECT y PET en los mejores resultados obtenidos con este método en la sección anterior. Dado que se desplaza arriba y a la izquierda, el método propuesto mejora otras técnicas publicadas como VAF, PCA en combinación con SVM [Angulo et al., 2008], [López et al., 2011] o el método continuo de ARs [Chaves et al., 2011a]. Se percibe que, dado que se seleccionan *minsup* y *minconf* al máximo de 100% (para los métodos basados en reglas tanto continuos como discretos), por consiguiente todos los controles se clasifican correctamente ($Spe = 1$ en el espacio ROC) dado que las ARs obtenidas son las más discriminantes para distinguir mejor los normales del patrón AD.

9.4 ARs como selector

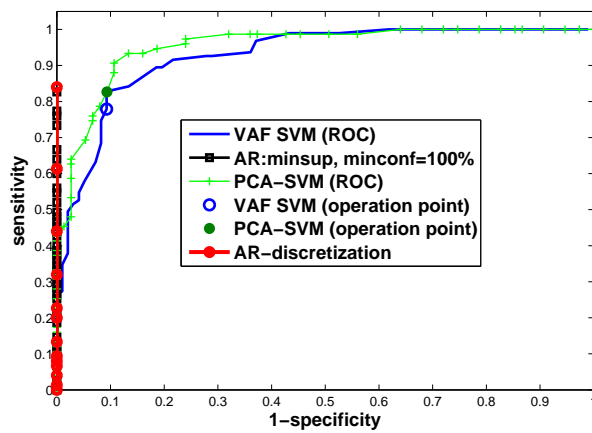
La sección 6.5 propone la idea de utilizar las ARs como selector de características con una posterior combinación con técnicas de extracción tales como PLS y PCA para su posterior clasificación con SVM. Los experimentos para las bases de datos SPECT y FDG PET son detallados a continuación.

9.4.1 Experimentos para SPECT

Para probar la fiabilidad del sistema CAD propuesto (cuyo diagrama explicativo aparece representado en la figura 6.9) llevamos a cabo algunos experi-



(a)



(b)

Figura 9.20: Curva ROC AR-discretos para bases de datos a)SPECT y b)PET. Comparación con el AR para características continuas y otros métodos publicados como: VAF-SVM y PCA-SVM.

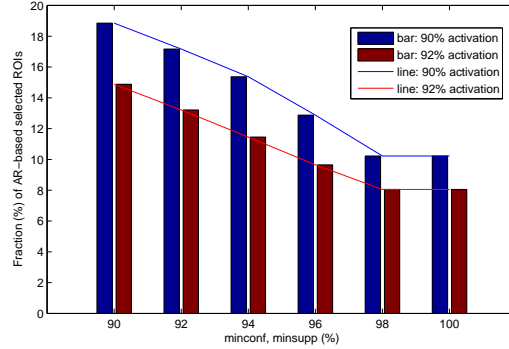
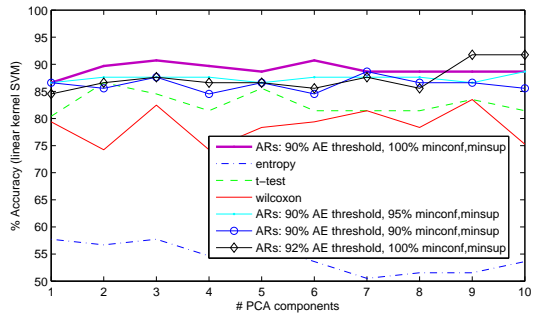


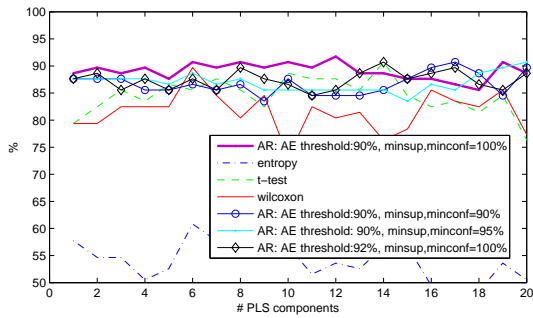
Figura 9.21: Fracción (%) de las ROIs seleccionadas mediante ARs frente a $minsup$ y $minconf$ (%) para diferentes parámetros: umbral de activación y $minsup/minconf$.

mentos. Después de la normalización espacial y en intensidad de la base de datos SPECT, se obtuvieron los bloques tridimensionales de tamaño $9 \times 9 \times 9$ localizados dentro de la máscara correspondiente. Las coordenadas centrales de los bloques 3D se restringen a un grid de tamaño $4 \times 4 \times 4$. En la fase de estimación de la activación, un bloque tridimensional se define como activado si más de una fracción 650 vóxeles de un total de 729 vóxeles están activados. Este umbral t representa un porcentaje por encima del 90% de los vóxeles activados, porque permite que sólo se consideren los volúmenes cerebrales con una alta activación como entrada para la extracción de ARs. En la fase de extracción de ARs, consideramos como entrada para el algoritmo *Apriori* diferentes valores de $minconf$ y $minsup$ (desde 100% a 90%). La mayoría de las reglas discriminativas se obtuvieron cuando se seleccionaron los valores más altos de $minsup$ y $minconf$. Además, el coste computacional se mejora reduciendo el número de características seleccionadas.

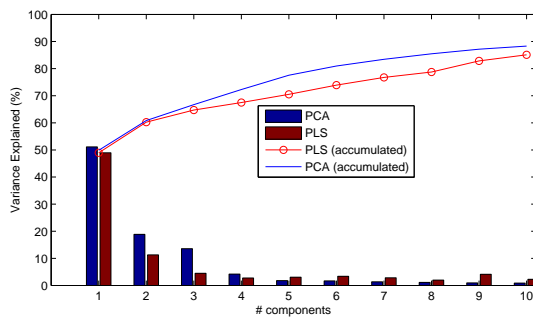
La figura 9.21 muestra la fracción (%) de las ROIs seleccionadas mediante ARs de un total de 2445 ROIs cuando se usan porcentajes de activación del 90% y del 92%. El número de reglas se reduce progresivamente de una forma similar a una distribución exponencial negativa desde el 90% al 100% de $minsup$ y $minconf$. Mientras que al 90% de $minconf$ y $minsup$, la fracción en % de la combinación de reglas seleccionadas del total de ROIs alcanza 18.85% y 14.88% (para activaciones del 90% y del 92%), al 100% este número aparece reducido al 10.22% o 8.05% respectivamente. A pesar de que se decrementa el número de ARs combinadas conforme el nivel de activación aumenta sobre el 90%, esto provoca la pérdida de reglas discriminativas para la detección



(a)



(b)



(c)

Figura 9.22: Valores de Acc en función del número de componentes (Base de datos SPECT, ARs como FS): a) PCA y b) PLS. c) Varianza explicada (%) como una función del número de componentes PCA y PLS en un diagrama de barras (las líneas representan la varianza explicada acumulada.)

	PCA: Acc/Sen/Spe (%)	PLS: Acc/Sen/Spe (%)	AUC: PCA/PLS
AR (90% AE, 100% <i>minconf</i> , <i>minsup</i>)	89.69/87.5/92.68	90.72/89.29/92.68	0.9549/0.9259
AR (90% AE, 95% <i>minconf</i> , <i>minsup</i>)	87.62/85.71/90.24	88.66/91.07/85.37	0.9017/0.9011
AR (90% AE, 90% <i>minconf</i> , <i>minsup</i>)	84.53/83.93/85.37	86.60/87.50/85.37	0.8915/0.8924
AR (92% AE, 100% <i>minconf</i> , <i>minsup</i>)	85.56/82.14/90.24	87.63/83.93/92.68	0.9277/0.9150
t-test	81.44/82.14/80.49	85.57/87.50/82.93	0.9317/0.9029
wilcoxon	79.38/80.36/78.05	89.69/91.07/87.80	0.9205/0.9230
energy	53.61/92.86/0	60.82/92.86/17.07	0/0

Tabla 9.5: Comparación de Acc/Sen/Spe y AUC para diferentes técnicas FE con reducción final de 6 características (SPECT)

temprana de la AD a la vista de los resultados; véase tabla 9.5 o figura 9.22(a) (para más detalles) cuando se utiliza un 92% de activación. Aunque esta reducción es considerable, una extracción de características con PCA o PLS es necesaria para mejorar el tiempo computacional y el ratio de exactitud del sistema. Los valores de Acc, Sen y Spe se estiman por medio de la validación cruzada *k-pliegues* con un clasificador lineal SVM. Las figuras 9.22(a) y 9.22(b) consideran un número creciente de características PCA o PLS respectivamente frente a la Acc (%) alcanzando valores de Acc, Sen y Spe máximas de 90.72%, 89.29%, 92.68% para 3 características PCA o 91.75%, 89.29%, 95.12% para 12 características PLS mejorando otros métodos recientemente publicados tales como t-student [Chaves et al., 2009b], wilcoxon [Bucchianico, 1996] o tests de entropía [Martínez et al., 2011]. El primero (t-student) es un t-test de dos muestras con estimación de varianza agrupada con pesado de correlación de características. El test de Mann-Whitney-Wilcoxon [Martínez-Murcia et al., 2012] y la entropía relativa nos da información acerca de la separabilidad de clase del vóxel pero con diferentes criterios de categoría. La reducción de dimensionalidad FE requiere una estrategia efectiva para encontrar el subconjunto óptimo de características o la transformación óptima para hacer una comparación fiable con otros métodos utilizados como partida. La figura 9.22(c) muestra el porcentaje de varianza explicada para las características extraídas dibujadas como barras y una línea representando la varianza acumulada explicada. La varianza explicada informa de la variación de un conjunto de características cuando se aplican las estrategias PCA o PLS y en este caso no cambia significativamente a partir de 6 componentes. La tabla 9.5 muestra la Acc, Sen y Spe del método de selección de características basado en ARs y la comparación con otras técnicas utilizadas como base en 6 características. Los valores de Acc, Sen, Sen convergen a 89.69%, 87.5% y

92.68% para PCA y 90.7%, 92.68% y 89.29% para PLS cuando se fija *minsup* y *minconf* a 100% y el umbral de activación a 90%, lo cual mejora los métodos de partida (t-test, wilcoxon o de entropía).

9.4.2 Experimentos para FDG-PET

Mientras que la intensidad del vóxel está relacionada con el flujo sanguíneo en las imágenes SPECT, para la modalidad de imagen PET el nivel de activación está relacionado con el metabolismo de la glucosa [Chaves et al., 2012b]. Hablando de forma genérica, las imágenes PET están menos activadas que las SPECT en cada ROI. Por esta razón los bloques tridimensionales se seleccionan en estos experimentos con un tamaño menor, es decir $5 \times 5 \times 5$ vóxeles por bloque, con coordenadas restringidas a un grid de $2 \times 2 \times 2$. La activación de un vóxel sobre el 40% permite incluir ROIs relevantes en la selección de características basadas en reglas.

La figura 9.23 muestra las características seleccionadas por el método de ARs propuesto en los cortes axial, sagital y coronal cuando *minconf* y *minsup* se seleccionan del 100% y 97% (primera y tercera fila respectivamente). La segunda y cuarta fila de la figura 9.23 resalta con una ampliación las regiones seleccionadas mediante las ARs. Es por ello que de una forma similar a una modalidad SPECT, el número de ROIs seleccionadas con reglas decrece más al 100% de *minconf* y *minsup* (de un total de 3776 ROIs), mejorando también el coste computacional. En particular, la fracción de ROIs seleccionadas basadas en ARs es 4.87% o 6.09% cuando la activación es del 44% o 40% respectivamente. Aunque al 44% de activación, la reducción es mayor, el sistema descarta algunas ROIs relevantes como se muestra en los resultados de exactitud de la figura 9.24.

Las Acc, Sen y Spe se estiman mediante la validación cruzada k -pliegues con un clasificador rbf SVM ya que opera mejor para la modalidad PET. Las figuras 9.24(a) y 9.24(b) consideraron un número creciente de características PCA o PLS respectivamente llegando a ratios máximos de Acc, Sen y Spe de 90%, 85.33%, 94.67% en 4 características PCA o 90%, 89.33%, 90.67% en 10 características PLS que mejoran otras técnicas publicadas como t-student, Mann-Whitney-Wilcoxon test y los basados en entropía relativa.

9.4.3 Análisis ROC

Las figuras 9.25 y 9.26 muestran las curvas ROC del sistema propuesto para las modalidades SPECT y PET considerando técnicas de extracción de características PCA (figuras 9.25(a) y 9.26(a)) y PLS (figuras 9.25(b) y 9.26(b)).

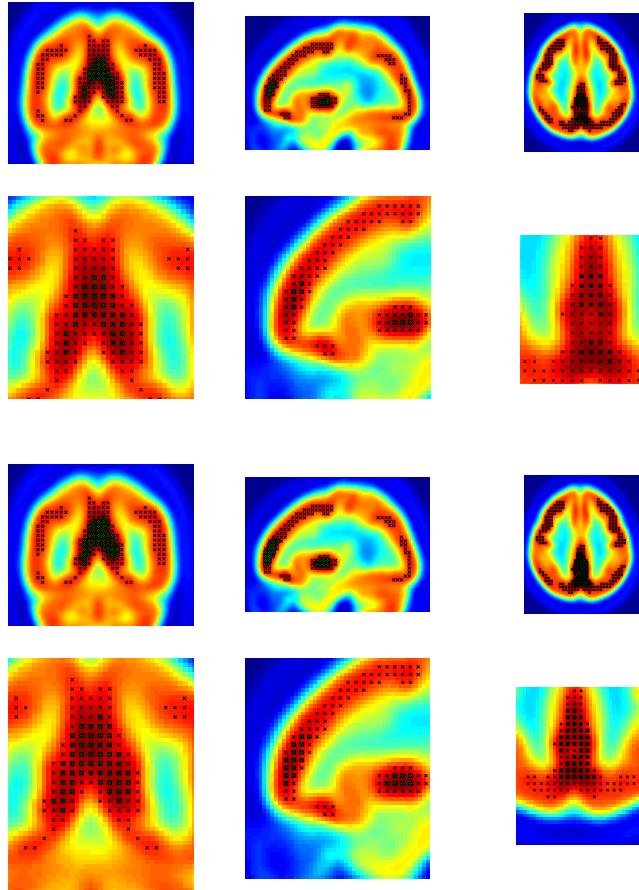
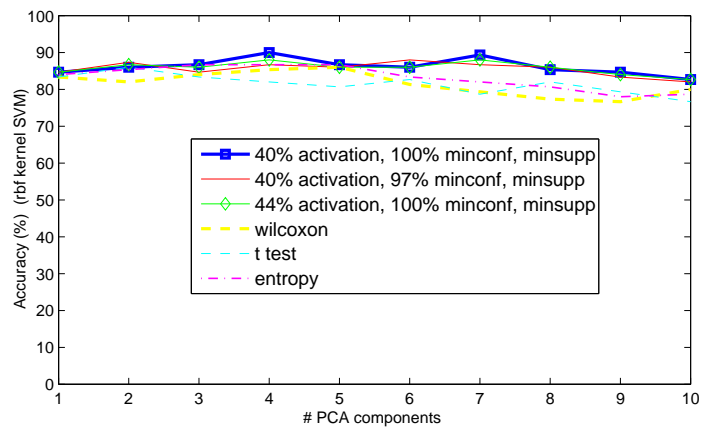
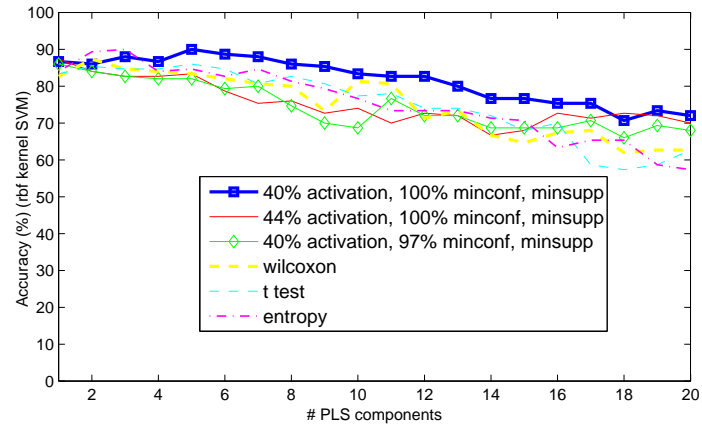


Figura 9.23: Vóxeles seleccionados para PET por el método de seleccion de reglas en los cortes axiales, sagitales y coronales. a) la primera fila con *minsup*, *minconf*=100%, b) la segunda fila amplía las regiones donde aparecen las ARs seleccionadas en a), c) tercera fila con *minsup*, *minconf*=97%, activación 40%, d) la cuarta fila amplía las regiones donde aparecen las ARs seleccionadas en c)

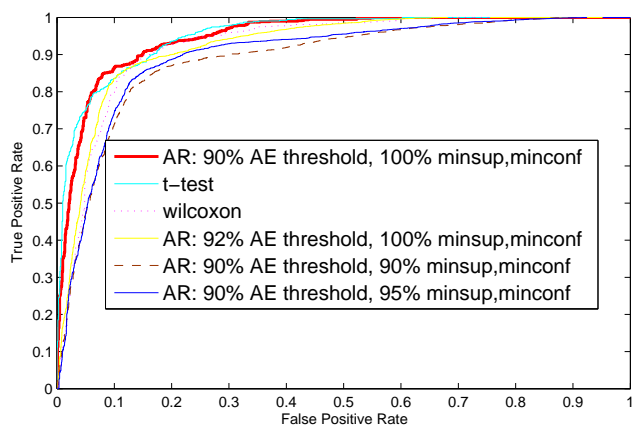


(a)

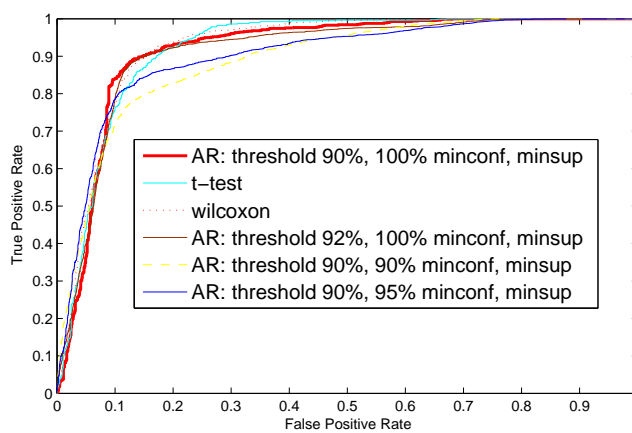


(b)

Figura 9.24: Valores de Acc función del número de componentes extraídas para la base de datos FDG-PET: a) PCA, and b) PLS. Método FS-AR.

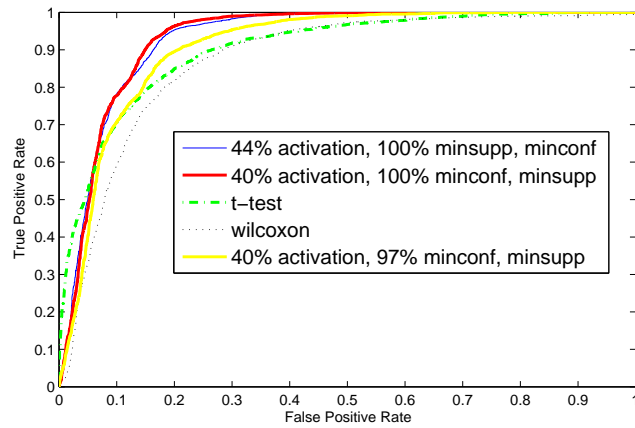


(a)

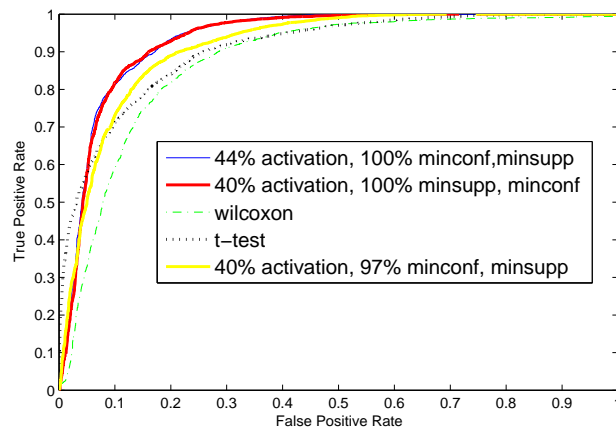


(b)

Figura 9.25: Curva ROC para la base de datos SPECT a)PCA y b)PLS. Método FS-AR. Comparación con otros métodos recientemente publicados.



(a)



(b)

Figura 9.26: Curva ROC PET a) PCA y b) PLS. Comparación con otros métodos recientemente publicados. La AUC obtenida para la ROC basada en AR-FS para PCA/PLS es: 100% *minconf* y *minsupp* con 40% activación (0.93441/0.9391), 44% activación (0.9107/0.9181) y 97% *minsupp* y *minconf* con 40% activación (0.9107/0.9181). AUC (PCA/PLS) para otros métodos publicados es: wilcoxon (0.8662/0.8662) o t test (0.8968/0.9027).

El sistema de selección basado en reglas mejora las técnicas base ya que su ROC se desplaza arriba y a la izquierda en el espacio ROC. Esta mejora del sistema propuesto se demostró también mediante el área bajo la curva (AUC), que está detallada en la tabla 9.5 para SPECT alcanzando valores de 0.9549 y 0.9259 para las modalidades PCA y PLS, respectivamente al 100% de *minconf* y *minsup* y 90% de umbral de activación. En lo que a PET se refiere, la AUC (detallada en la leyenda de la figura 9.26) muestra una tendencia similar a SPECT, es decir llega a los resultados más altos de 0.9344 y 0.9301 para las modalidades PCA y PLS al 100% de *minconf* y *minsup* y 40% de umbral de activación.

Capítulo 10

Resultados experimentales con LMNN

En este capítulo mediante una serie de experimentos aplicados en imagen funcional, hemos investigado la habilidad del algoritmo LMNN para evaluar (con k -pliegues) los algoritmos de extracción de características y sus beneficios como: i) transformación lineal de los datos previamente reducidos con PLS o PCA, ii) técnica de reducción de características, y iii) clasificador basado en distancias euclídea, de mahalanobis o bien en modelos basados en energía. La clasificación SVM incluyendo la transformación del espacio de entrada en término de los “kernels” (lineal, polinomial, cuadrático o rbf) que definen las superficies de decisión no lineales se adoptaron para las dos primeras aproximaciones (i e ii) . La selección del modelo y las características se lleva a cabo en base a lo explicado en el capítulo 7 con las bases de datos SPECT y FDG-PET de ADNI. También se incluyen los experimentos obtenidos mediante la transformación no lineal lograda mediante el truco del kernel que simplifica la transformación de LMNN en kernel LMNN (KLMNN).

La alternativa del algoritmo LMNN consiste en construir una distancia de Mahalanobis sobre el espacio de entrada y usarla en lugar de las distancias euclídeas. Esta aproximación puede ser interpretada equivalentemente como una transformación lineal de las entradas originales, seguidas por la distancia euclídea en el espacio proyectado. También se incluyen en este capítulo los experimentos obtenidos mediante la transformación no lineal que se logra mediante el truco del kernel.

10.1 Selección del modelo y las características

Las fases del sistema CAD basado en el algoritmo LMNN se detallaron en la sección 7.2 y se resumen en:

- Enmascarado
- Extracción de características: se aplica un t -test a las ROIs tridimensionales con valor su NMSE. La reducción posterior se realiza con PCA o PLS transformadas con LMNN o bien mediante LMNN-RECT.
- Clasificación con algoritmo LMNN usando distancia euclídea, de Mahalanobis o con métodos basados en energía o bien mediante kernel SVM.

En el proceso de reducción de características, hay ciertos parámetros para ajustar tales como en número de bloques NMSE, el número de características reducidas PCA, PLS o LMNN-RECT y la selección del kernel (lineal, polinomial, cuadrática o RBF) que define mejores superficies de decisión en la clasificación SVM. Las características NMSE se calcularon usando bloques de vóxeles de $5 \times 5 \times 5$ ya que tamaños cúbicos reducidos de características NMSE logran mejores resultados como se muestra en [Chaves et al., 2009b]. Además, se seleccionaron 200 características discriminativas en términos de la reducción t -test (un número mayor de bloques NMSE significa un decrecimiento de la efectividad del método de clasificación). La reducción posterior del tamaño del vector de características se lleva a cabo en términos de PCA, PLS o LMNN-RECT.

El número de características usadas es un compromiso entre las ROIs que son realmente importantes y que no empeoran el tiempo computacional del CAD. Demostramos experimentalmente que 200 características NMSE (ROIs) es un número lo suficientemente alto para garantizar la calidad de la imagen en la clasificación posterior. Sin embargo, este número debe ser reducido para mejorar el tiempo computacional del sistema con estrategias como PCA, PLS

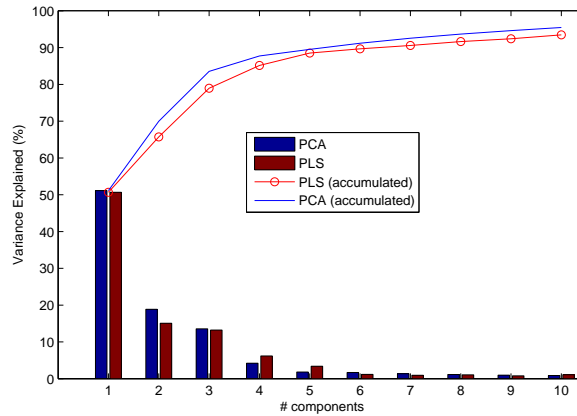


Figura 10.1: Varianza explicada (%) frente a las componentes principales y PLS en diagrama de barras. Las líneas representan la varianza explicada acumulada (%) frente a las componentes principales y las componentes PLS.

o LMNN-RECT. El número final de características usado ha sido experimentalmente elegido como se observa en la figura 10.1 en la que el porcentaje de la varianza explicada para las características (PCA o PLS) elegidas se dibujan como barras y una línea representa la varianza explicada acumulada. En este caso, la varianza explicada justifica la elección del conjunto de características cuando se aplican las estrategias PCA o PLS. En este gráfico, podemos observar que por encima de 6 componentes, la varianza explicada para PCA y PLS no varía de forma significativa [Chaves et al., 2012d].

10.2 Evaluación mediante *k* pliegues

El rendimiento de la clasificación de nuestra aproximación se probó en términos de la validación cruzada *k*-pliegues (como se explica en detalle en la sección 5.6), que es bastante utilizada hoy en día para comparar los rendimientos de diferentes procedimientos de modelado predictivo como en [Wiens et al., 2008]. Aunque hay estudios que consideran *k* separaciones independientes de entrenamiento y test (por ejemplo en [Nadeau and Bengio, 2003] y [Bengio and Grandvalet, 2004]), nos centramos en el estándar de la validación cruzada en *k*-pliegues como en [Westman et al., 2011] y [Bengio and Grandvalet, 2004]. En el procedimiento de los *k*-pliegues, no hay solapamiento entre los conjuntos de test: cada ejemplo del conjunto original de datos se usa sólo una vez como

un ejemplo de test. A menudo llamada estimación por rotación, el conjunto de datos D se separa aleatoriamente en k subconjuntos mutuamente exclusivos (los pliegues) D_1, D_2, \dots, D_k de aproximadamente igual tamaño. El inductor se entrena y testea N veces; cada vez $t \in \{t_1, t_2, \dots, t_N\}$, se entrena en $D - \{D_t\}$ y se testea en D_t [Kohavi, 1995]. En este capítulo se usaron 10 pliegues en cada experimento que alcanzaron estimaciones exactas de las tasas de error. Para cada iteración ($t=1, \dots, 10$), el algoritmo devuelve aleatoriamente los índices generados para la validación cruzada de k -pliegues de D observaciones. La tasa de test es mayoritariamente igual al entero de la fracción $100/\text{número de pliegues}$, es decir el 10% de los experimentos, pero puede variar aleatoriamente una o dos muestras en cada iteración si el número de observaciones es un número primo. Estos índices se usan para testear y el resto (aproximadamente el 90%) para entrenar. Los resultados estadísticos obtenidos en cada iteración se promedian.

Por lo tanto, usando validación cruzada, algunos métodos de extracción de características y clasificación se compararon objetivamente en términos de sus fracciones respectivas de muestras mal clasificadas. En este sentido, el clasificador se evalúa en profundidad como una herramienta para la detección temprana de la AD en términos de la Acc, Sen y Spe. Para un análisis posterior, los datos se organizan en dos grupos diferentes: los sujetos AD fueron etiquetados como positivos y los controles como negativos. La motivación de hacer esto es testear nuestro método con todas las fases disponibles de la enfermedad, manteniendo la base de datos tan balanceada como sea posible (41 CTRL frente a 56 AD para SPECT y 75 CTRL frente a 75 AD para ADNI FDG-PET) e incluir varios tipos de patrones en la tarea de clasificación (entrenamiento y test).

10.3 Experimentos con la base de datos SPECT

Apéndice de figuras (SPECT):

1. Figura 10.2: Clasificación SVM; Acc, Spe y Sen (%) frente al número de características reducidas mediante a) PCA+Transformación LMNN, b)PLS+Transformación LMNN, c)LMNN-RECT
2. Figura 10.3: Clasificación LMNN con distancias euclídea, de Mahalanobis y con modelos basados en energía. Técnicas de FE: a)PCA, b)PLS
3. Figura 10.4: a) Kernel SVM para características PLS transformadas con

LMNN. b) SVM lineal sobre características directas PCA reducidas a la mitad.

En un primer experimento, se compararon las técnicas propuestas en este capítulo (PCA/PLS+ Transformación LMNN y LMNN-RECT) en las figuras 10.1(a-c). Se demuestra que todos los métodos de extracción de características son muy robustos para la selección del número de características de entrada en el clasificador por la estabilidad mostrada en las gráficas. Cuando se usan 6 características en clasificación, PCA-LMNN alcanza una $Acc=91.75\%$, $Sen=91.07\%$ y $Spe=92.68\%$, mientras que PLS-LMNN mejora estos valores con una $Acc=92.78\%$, $Sen=91.07\%$ y $Spe=95.12\%$. Referente a LMNN-RECT, se obtienen los mejores resultados: $Acc=90.72\%$, $Sen=96.43\%$ y $Spe=82.93\%$, con 18 características. La ventaja de LMNN-RECT es su simplicidad y rapidez porque no necesita la combinación con otra técnica de reducción ni tampoco una transformación LMNN.

En el segundo experimento se analiza la clasificación usando LMNN con las distancias Euclídea, de Mahalanobis y los modelos basados en energía cuando se consideran las características NMSE extraídas con PCA o PLS. Las figuras 10.2.a y 10.2.b muestran la exactitud obtenida mediante la clasificación LMNN como una función del número de coeficientes PCA y PLS, respectivamente. Los resultados muestran que la clasificación LMNN usando modelos basados en energía y distancias de Mahalanobis consiguen mejores resultados cuando se considera la distancia Euclídea, que sufre un decremento en la exactitud conforme el número de características crece. Se demuestra que tanto la clasificación LMNN usando los modelos basados en energía como las distancias de Mahalanobis son modelos robustos frente a la dimensión del vector de características alcanzando valores de pico en la exactitud de 91.78% y 89.78% para modelos basados en energía y distancia de Mahalanobis respectivamente. En todos los casos, los modelos basados en energía mejoraron los otros.

Dado que con la extracción de características mediante PCA y PLS en combinación con una transformación LMNN conseguimos los mejores resultados y con el objetivo de mejorar la exactitud de la clasificación, analizamos la selección del kernel SVM más adecuado. En efecto, la figura 10.4.a muestra la exactitud del sistema como una función del número de coeficientes PLS para la clasificación SVM basada en los kernels lineal, polinómico y RBF. En conclusión, el kernel lineal obtiene mejores resultados de exactitud con un 92.7% . Es significativo que cuando se combinan los bloques 3D NMSE como características de entrada y después se transforman con el algoritmo LMNN en sus múltiples posibilidades (como técnica de reducción, transformador lineal o clasificador) se añade robustez al sistema. Esto se ha demostrado con los ex-

clasificador SVM lineal	Acc (%)	Sen (%)	Spe (%)
VAF	83.51	83.93	82.93
PCA	86.56	91.07	80.49
GMM	89.69	90.24	89.29
kernel gaussiano PCA + transformación LMNN	91.75	91.07	92.68
kernel gaussiano PLS + transformación LMNN	90.72	91.07	90.24
PLS + transformación LMNN	92.78	91.07	95.12
LMNN-RECT	80.28	70	87.80
Exactitud del clasificador LMNN (%)	Euclídea	Mahalanobis	Energía
PCA	80.54	81.63	87.65
PLS	84.33	89.56	88.67

Tabla 10.1: Medidas estadísticas de rendimiento de las técnicas basadas en LMNN en comparación con otros métodos publicados para la base de datos SPECT. Clasificador SVM: Comparación de los métodos publicados en este trabajo con los puntos de operación de VAF, GMM y PCA. Parámetros de las técnicas basadas en LMNN: SVM lineal con 6 componentes. Parámetros VAF: clasificador SVM lineal, parámetros GMM: clasificador RBF-SVM con $\sigma=6$, parámetros PCA: clasificador RBF-SVM con 16 componentes. Clasificador LMNN con 6 componentes euclídeas, de Mahalanobis y basada en energía.

perimentos de las figuras 10.1, 10.2, 10.3 y 10.4.a. Como contraste en la figura 10.4.b, PCA se usa directamente sobre los vóxeles reducidos a la mitad (para aliviar el coste computacional) y tratados con el mismo tipo de máscara que en este capítulo. Los resultados de la figura 10.4.b mostraron que la variación de la exactitud aumenta cuando los vóxeles se utilizan como características, haciendo el sistema menos robusto. Sin embargo, en este capítulo se demuestra que la ventaja de la combinación de los métodos propuestos basados en LMNN es que se mantienen estables alrededor del 90%. Podemos concluir que el hecho de escoger las ROIs usando la combinación de la técnica NMSE con el algoritmo LMNN favorece la estabilidad en todo el rango de características producidas, promoviendo por tanto la robustez del algoritmo.

Finalmente la tabla 10.1 muestra los ratios de Acc, Sen, Spe de los métodos propuestos y los compara con otros recientemente publicados incluyendo VAF, GMM [Segovia et al., 2010] y PCA [Stoeckel et al., 2001], [López et al., 2009a]. Basándonos en el modelo de reducción de características mostrado en la sección 7.2.1, es decir mediante kernel PCA y PLS, y como se muestra en la figura

10.1 se demuestra que hasta 6 características retienen la mayor parte de la varianza de los datos. En la tabla 10.1 para 6 componentes se demuestra que la transformación LMNN cuando se combina con PCA o PLS logra los mejores resultados y reporta beneficios cuando se compara con los otros métodos de referencia. Además LMNN se presentó como una solución válida para aumentar el margen entre clases. Hemos desarrollado un sistema CAD efectivo en el que no es necesario incorporar un conocimiento *a priori* de la patología, ya que a partir del primer paso de la extracción de características, se consideran todos los vóxeles con una activación considerable (es decir, aquellos vóxeles que se localizan en la máscara calculada). El análisis mostrado en este capítulo informa también de las claras ventajas en los consecutivos pasos de selección de ROIs, dado que fueron calculados de forma automática para el diagnóstico precoz de la AD. La mejor combinación de las técnicas de reducción de características alcanza (véase tabla 10.1 un valor de exactitud de 92.72%, por lo tanto mejorando otros métodos recientemente publicados y consolidados como VAF, PCA o GMM).

Finalmente, con objeto de estudiar en profundidad la clasificación de AD con técnicas basadas en LMNN, hemos incluido información adicional de la clasificación de AD1 frente a CTRL. Este paso es más difícil de clasificar ya que el patrón AD1 tiene un diagnóstico precoz que a día de hoy constituye un reto. Si sólo consideramos el caso CTRL frente a AD1, los ratios de precisión son para PCA más la transformación LMNN: Acc =84.51%, Sen=73.33%, Spe=92.68%, para PLS más la transformación LMNN: Acc=83.10%, Sen=70%, Spe=92.68% y para LMNN-RECT: Acc=84.51%, Sen=76.67%, Spe=90.24%. Estos resultados aún representan un gran avance en el campo en comparación con VAF: Acc=77.46%, Sen=70%, Spe=82.93%.

10.4 Experimentos con la base de datos PET

Apéndice de figuras (ADNI-FDG PET):

1. Figura 10.5.a: Clasificación SVM; Acc, Spe y Sen (%) frente al número de características reducidas mediante a) PCA+Transformación LMNN, b)PLS+Transformación LMNN, c)LMNN-RECT y comparativas.
2. Figura 10.5.b: Clasificación LMNN con distancias euclídea, de Mahalanobis y con modelos basados en energía. Técnicas de FE: a)PCA, b)PLS

Adicionalmente se llevaron a cabo experimentos en la base de datos PET para resaltar la habilidad de generalización del método propuesto. Para ello

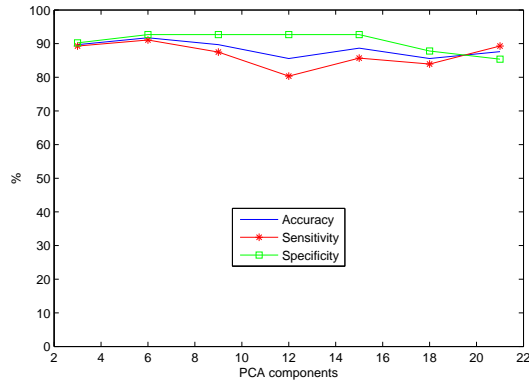
se utilizaron los mismos parámetros que en SPECT tales como el tamaño de vóxel o el número de bloques NMSE. Como ejemplo la figura 10.5.a muestra las diferentes técnicas de extracción de características de este trabajo, es decir, PCA o PLS más la transformación LMNN y LMNN-RECT en comparación con PCA o VAF cuando se usa un clasificador lineal SVM. Se alcanza una Acc de 90.67% (88% Sen y 93.33% Spe) para tanto PCA como PLS más la transformación LMNN y cuando se usa LMNN-RECT, la exactitud alcanzada es de 87.33% (82.67% Sen y 92% Spe), por lo tanto se mejoran PCA (Acc=85.33%) o VAF (Acc=81.18%). La figura 10.5.b muestra la clasificación LMNN usando modelos basados en energía, distancias de Mahalanobis o Euclídeas para las características PCA o PLS. Los ratios máximos de exactitud los obtuvimos para el clasificador basado en energía (90.11% para PCA y 89.99% para PLS).

10.5 Análisis ROC

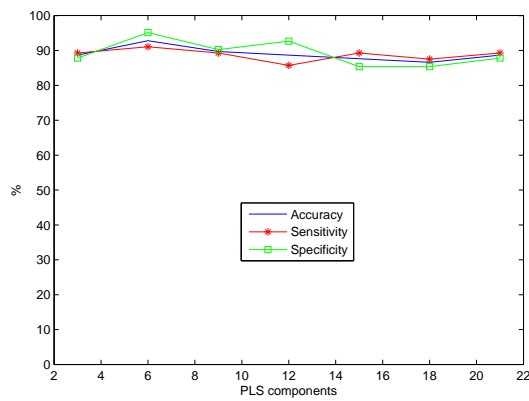
Apéndice de curvas ROC para SPECT y FDG-PET

1. Figura 10.6.a: Curva ROC SPECT para PCA+Transformación LMNN, PLS+ Transformación LMNN, LMNN-RECT, VAF SVM y PCA SVM.
2. Figura 10.6.b: Curva ROC ADNI-FDG PET para PCA+ Transformación LMNN, PLS+ Transformación LMNN, LMNN-RECT, VAF SVM y PCA SVM.

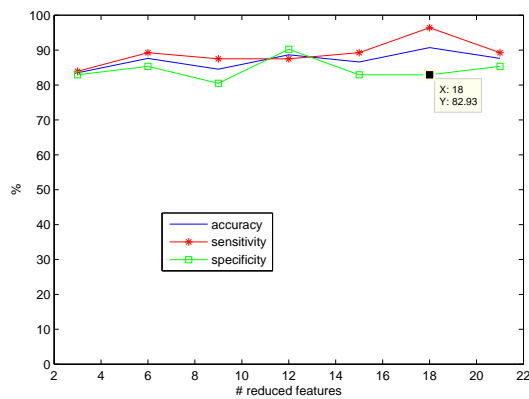
Las figuras 10.6(a-b) muestran las curvas ROC (véase información teórica detallada en la sección 5.6.1) de los sistemas propuestos y otros métodos que se consideraron como referencia para las bases de datos SPECT y PET respectivamente. El análisis muestra que el sistema CAD basado en el algoritmo LMNN y SVM consigue el mejor compromiso entre la sensibilidad y la especificidad ya que desplaza el punto de operación arriba y a la izquierda en el espacio ROC [Metz, 1978] en comparación con otros métodos publicados como VAF SVM, PCA SVM y GMM SVM. Como se muestra en la figura 10.6.b, PLS más la transformación LMNN provee un punto de operación localizado en la parte más superior y a la izquierda respecto a otras aproximaciones en el espacio ROC. Además, la mejora de la técnica basada en LMNN se apoya también por el análisis AUC para SPECT/PET respectivamente: PLS más la transformación LMNN: 0.9424/0.9437, PCA más la transformación LMNN 0.9411/0.9505, LMNN-RECT: 0.9076/0.9325 que mejoran la AUC de otros métodos publicados tales como VAF SVM: 0.8993/0.8500 y PCA SVM: 0.9177/0.9006.



(a) PCA+ transformación LMNN

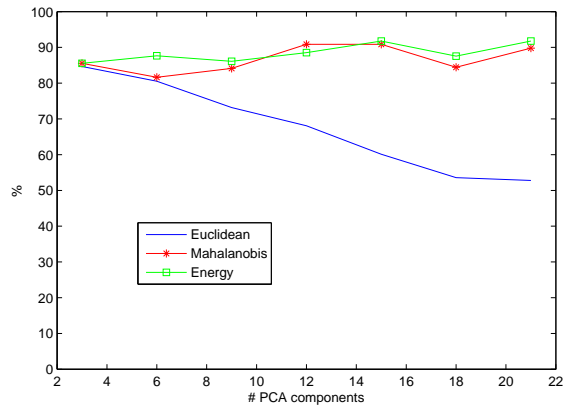


(b) PLS+ transformación LMNN

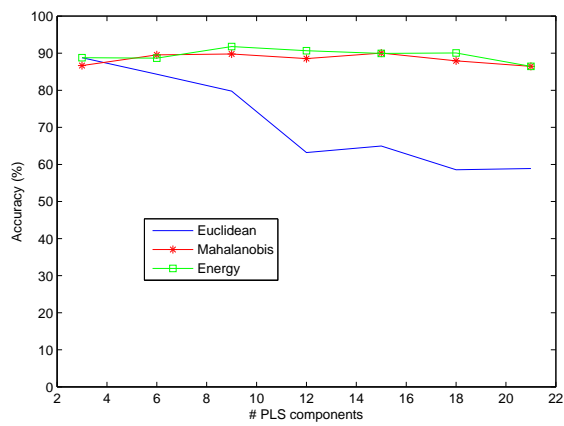


(c) LMNN-RECT

Figura 10.2: Clasificación SVM: Acc, Spe y Sen (%) frente al número de características reducidas para SPECT. Técnicas de reducción de características a) PCA + Transformación LMNN, b) PLS + Transformación LMNN, c) LMNN-RECT

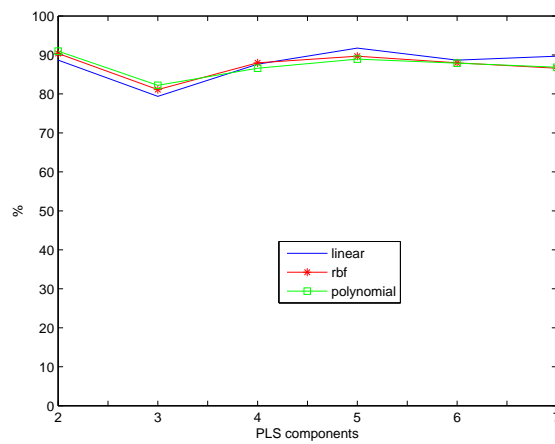


(a) PCA

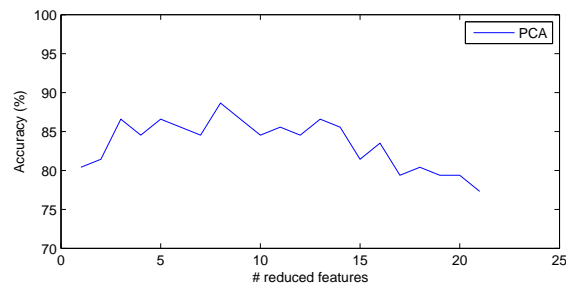


(b) PLS

Figura 10.3: Clasificación LMNN (Euclidea, Mahalanobis y modelos basados en energía) para SPECT. Técnicas de reducción de características a)PCA, b)PLS

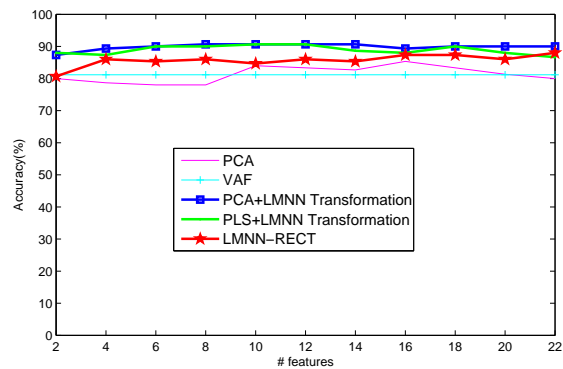


(a) Kernel SVM de características transformadas PLS+LMNN

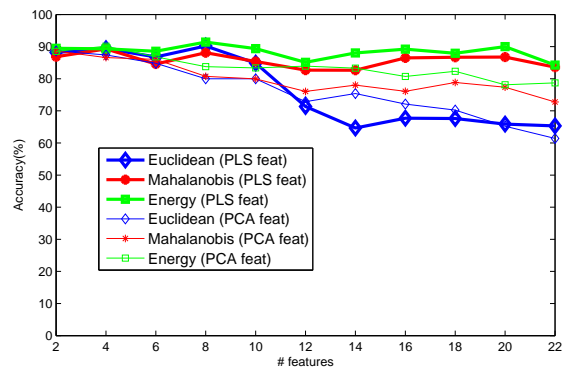


(b) PCA

Figura 10.4: a) Kernel SVM para características PLS transformadas LMNN
b) SVM lineal sobre características directas PCA reducidas a la mitad VAF obtenidas para SPECT

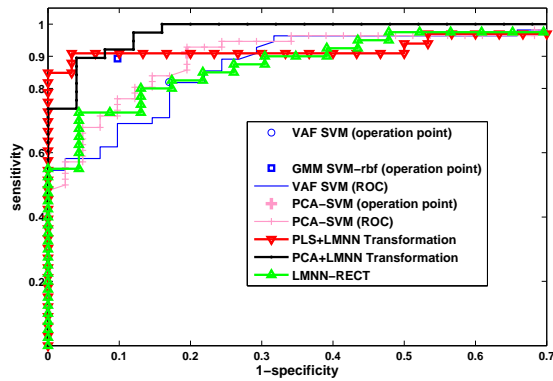


(a) Clasificador SVM

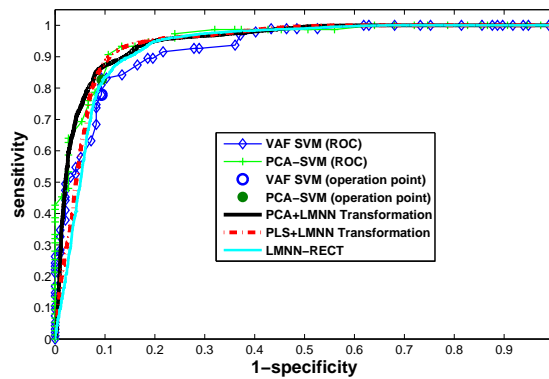


(b) Clasificador basado en LMNN

Figura 10.5: a) Clasificación SVM Acc, Spe y Sen (%) frente al número de características reducidas para la base de datos PET. Técnicas de reducción de características: PCA más transformación LMNN, PLS más transformación LMNN, LMNN-RECT, PCA y VAF b) Clasificación LMNN (euclídea, mahalanobis y modelos basados en energía) para PCA y PLS



(a) Base de datos SPECT



(b) Base de datos

Figura 10.6: Los métodos basados en LMNN clasificados con SVM (Transformación PCA+ Transformación LMNN, PLS+Transformación LMNN, LMNN-RECT). Comparación con otros métodos recientemente publicados representados por sus puntos de operación. a) Base de datos SPECT. AUC: PCA+LMNN (0.9411), PLS+LMNN (0.9424), LMNN-RECT (0.9076), VAF SVM (0.8993) y PCA SVM (0.9177). b) PET database. La AUC obtenida para cada ROC es: PCA+LMNN (0.9505), PLS+LMNN (0.9437), LMNN-RECT (0.9325), VAF SVM (0.8500) and PCA SVM (0.9006).

Parte VI

Conclusiones y discusión

Capítulo 11

Conclusiones y trabajo futuro

Este capítulo recoge las conclusiones y aportaciones científicas de esta Tesis doctoral para el diseño de sistemas CAD que de forma efectiva permiten la detección precoz de alteraciones neurológicas como la que nos ocupa en este trabajo, la AD. La investigación realizada se organiza en dos vertientes: ARs y LMNN. En primer lugar se proponen las ARs como una técnica novedosa y eficaz para el análisis de imagen funcional cerebral entre atributos continuos, discriminantes y/o discretos, cuya clasificación se lleva a cabo mediante un sistema basado en el número de ARs verificadas por cada paciente. Además las ARs se aplican como selector de características que se combinan posteriormente con distintas técnicas de extracción de características (tales como PCA y PLS) y se clasifican con “kernel” SVM. En segundo lugar, dentro del problema de la métrica de distancias, proponemos el algoritmo LMNN como extractor y transformador del espacio de características así como clasificador basado en las distancias Euclídea, de Mahalanobis o en métodos basados en energía. Por último, tras la evaluación de los métodos propuestos se presentan algunas posibilidades de trabajo futuro para la continuación y mejora de las líneas de investigación iniciadas.

11.1 Discusión y conclusiones de este trabajo

La motivación y aportación principal de este trabajo radicó en el diseño y experimentación de sistemas CAD completos que discriminaron entre patrones normales y de hipoperfusión e hipometabolismo con la finalidad de diagnosticar precozmente una alteración neurológica como la AD. Los algoritmos desarrollados fueron evaluados mediante las diferentes bases de datos descritas en el capítulo 8 que son: la base de datos SPECT de 97 imágenes (Hospital Virgen de las Nieves, Granada), FDG-PET de 150 imágenes (procedente de ADNI), 103 pacientes adquiridos por separado con los biomarcadores FDG y PIB (de ADNI) y 60 imágenes FDG-PET procedentes de la clínica privada PET Cartuja (Sevilla). Las bases de datos utilizadas no están patológicamente confirmadas y por consiguiente pueden introducir cierta incertidumbre en las etiquetas de los sujetos, dado que es posible que el patrón de perfusión no refleje la patología de la AD subyacente, ni la severidad de los síntomas de los pacientes. Como contraste cuando la información patológica está disponible, se demostró que la evaluación visual realizada por expertos es sensible y específica para el etiquetado, en contraste con los tests neuropsicológicos [Dubois et al., 2007], [Jobst et al., 1998]. En esta misma línea, el hecho de usar estas etiquetas nos permitió probar la robustez de los diferentes clasificadores así como comparar entre ellos y con otros métodos testeados en pacientes AD con autopsia confirmada de los que se esperaba que cada clasificador mejorase su rendimiento. Así se demostró que este trabajo podría usarse no sólo para la valoración cuantitativa sino también para valoración visual en la práctica clínica. Sin embargo, la interpretación cuantitativa de los datos de precisión alcanzados debe realizarse con cautela ya que existen diferencias de etiquetado entre bases de datos. Por ejemplo, en PET-Cartuja se tiene en cuenta la valiosa información de las imágenes para el etiquetado mientras que en PET-ADNI sólo se consideran los tests. A continuación se exponen las conclusiones de este trabajo en las dos líneas de análisis de imagen funcional propuestas en la parte 4 de este trabajo, es decir ARs y LMNN.

11.1.1 Reglas de asociación

Hasta el momento, las ARs principalmente se han utilizado en economía para el análisis de mercados [Agrawal et al., 1993]. La habilidad de esta técnica para establecer relaciones entre datos no sólo tiene utilidad para asociar el consumo de ciertos productos sino que puede tener una aplicación directa en neuroimagen para interconectar regiones cerebrales activadas. Esta ha sido nuestra principal motivación en este trabajo, en el que enfocamos el problema

desde diferentes perspectivas (véase capítulo 6): ARs obtenidas entre atributos continuos, entre atributos discriminantes para mitigar el coste computacional que las ARs acarrearán y por último entre atributos discretizados para evitar el uso de umbrales de activación al considerar si una región está activada (o no). Para estos casos la clasificación se llevó a cabo según el número de reglas verificadas siguiendo la evaluación LOO (véase sección 5.6). Por otro lado, se presentó además un modelo eficaz de selección de características basado en ARs combinado con métodos de extracción de características (tales como PCA y PLS) y clasificación “kernel”-SVM.

ARs entre atributos continuos

En primer lugar, investigamos las ARs entre atributos continuos para la clasificación de imágenes tomográficas y para el diseño de un sistema CAD con la finalidad de diagnosticar precozmente la AD. Los atributos continuos son frecuentes en la naturaleza y su comprensión se hace indispensable para el diseño de un sistema de minería de datos. Dado que un atributo continuo puede adquirir multitud de valores (en nuestro caso valores de activación), tuvimos que definir un umbral para determinar si una región tridimensional resultaba interesante para interconectar el cerebro mediante ARs o no como detalló la sección 6.2. Tras la fase de enmascaramiento (VAF, descrita en la sección 6.2.1) obtuvimos las regiones tridimensionales centradas en las coordenadas de la máscara, a lo que le siguió la fase de estimación de la activación (AE) que fijaba un umbral $t = 90\%$ dado que las ROIs altamente activadas nos ayudaban a detectar mejor la AD como se explicó en la sección experimental 9.1. El sistema propuesto se basó en la extracción de ARs de los sujetos controles y en su clasificación, dependiendo del número de reglas verificadas por cada paciente. Además, este sistema CAD se definió como un método “offline”, dado que las ARs fueron extraídas mediante el algoritmo *A priori* sólo a partir del conjunto de controles (dado que definen un patrón más estable que el AD) y cada sujeto de test se evaluó en términos del mismo conjunto de reglas. Es importante resaltar que las ARs se extrajeron en términos de la estrategia de validación cruzada LOO [Gidskehaug et al., 2008], es decir si se evaluaba un control determinado, no se tenía en cuenta en el proceso de extracción de ARs para de esta forma evitar el sobreentrenamiento. La clasificación de cada paciente se basó en el número de ARs que verificaba y se demostró que el rendimiento del CAD en términos de su Acc, Spe y Sen se iba incrementando al valor máximo conforme el umbral de reglas verificadas aumentaba. Sin embargo, esta tendencia sólo ocurría cuando las ARs se extraían como las más discriminativas, con valores de *minsup* y *minconf* del 100%, en otro caso los ratios disminuían

a un número determinado de reglas, incluso cuando se elegían las ARs más adecuadas en clasificación. En particular los mejores resultados obtenidos en términos de Acc, Spe y Sen fueron de: 94.85/91.07/100 % para SPECT y de 93.33/90.48/100 % para la base de datos PET Cartuja.

Aparte de analizar experimentalmente este método usando las bases de datos SPECT y PET-Cartuja, con el mismo tamaño de vóxel y solapamiento, en la sección 9.1.3 se mostraron también los resultados de clasificación al estudiar por separado o combinados los biomarcadores *FDG* y *PIB* de la base de datos ADNI de 103 pacientes. Esta idea surgió de artículos como [Davatzikos et al., 2010] en los que se demostró una mejora en clasificación al utilizar información complementaria de los biomarcadores SPARE-AD y CSF. La comparación se llevó a cabo en los conjuntos CTRL frente a AD (conjunto 1) y CTRL frente a MCI (conjunto 2). PIB se distinguió de FDG en que proveía información diferencial para el diagnóstico precoz de la AD. A la vista de los resultados mostrados en el capítulo 9, concluimos que la combinación de los biomarcadores en el primer conjunto mejora los resultados de FDG mientras que en el segundo conjunto se mejoran ambos ratios (para FDG y PIB). En el conjunto 1, con el biomarcador PIB se obtuvieron mejores resultados *Acc/Sen/Spe* de 92.11/84.21/100 %, mientras que en el conjunto 2, la combinación de FDG y PIB aportó mayor calidad al sistema con ratios de clasificación 92.85/90.77/100 %. En definitiva con este análisis deducimos que la ventaja de los métodos basados en ARs respecto a otros métodos propuestos usados en la comparativa (tales como PCA, VAF etc) es que mientras estos son estudiados desde un punto de vista estático, usualmente fijándose en los patrones de hipoperfusión e hipometabolismo, el método basado en ARs es innovador ya que estudia las relaciones entre áreas cerebrales con ARs aportando dinamismo al sistema CAD. Todos los métodos propuestos en esta Tesis fueron cotejados con las comparativas mediante el análisis de curvas ROC y se demostró que al estar desplazados hacia arriba a la izquierda en el espacio ROC [Metz, 1978], los resultados mejoraban sensiblemente.

ARs entre atributos discriminativos

En segundo lugar, estudiamos las ARs entre atributos discriminativos dado que la principal problemática de las ARs radicaba en el coste computacional, tema aliviado mediante el uso de máscaras basadas en un umbral *FDR* como se explicó en la sección 6.3. Una máscara FDR no sólo permitió seleccionar las regiones más discriminativas para un análisis y mejora de los resultados de clasificación, sino que también contribuyó a la reducción del coste computacional, obteniéndose resultados optimizados con un umbral FDR $h = 0.2$, como

se mostró en la sección 9.2. Las ARs se continuaron extrayendo del patrón normal y la regla de decisión se basó en el número de reglas verificadas por cada sujeto. Además usamos dos bases de datos SPECT y FDG-PET ADNI con los mismos parámetros (tamaño de vóxel y grid) para resaltar la habilidad de generalización del sistema CAD aunque con un tamaño de vóxel y umbral de activación reducidos ($t=40\%$) respecto a los experimentos del caso continuo, dado que así lo forzaba la base de datos PET ADNI porque presentaba regiones menos activadas respecto a SPECT. El método propuesto mejoró otros recientemente publicados (como VAF o PCA) convergiendo a un punto de operación óptimo cuando el número de ARs verificadas es del 100% y las ARs se extrajeron al 100% del *minconf/minsup*, es decir eran las más discriminativas. En definitiva, para SPECT alcanzamos ratios *Acc/Sen/Spe* máximos de 92.69/87.5/100 %, mientras que para FDG-PET fueron de 91.33/82.67/100 %.

ARs entre atributos discretos

En tercer lugar, se describieron las ventajas de combinar atributos discretos y relacionarlos mediante ARs en bases de datos FDG-PET y SPECT. Discretizar evitó (respecto al caso continuo) la necesidad de utilizar un umbral t para determinar si una característica era de interés o no. De esta forma la selección de las ROIs entre las que se establecieron las ARs resultaba un proceso más automatizado. Como se explicó en la sección 9.3 primero se aplicó una segmentación basada en histograma de imagen sobre la imagen media de controles para obtener la mejor máscara correspondiente a los intervalos más altos (en particular el cuarto para SPECT y quinto para PET, de un total de K_1 subimágenes). La selección de $K_1 = 5$ para la obtención de la máscara implicó una generalización en ambas bases de datos y el hecho de asumir que el histograma basado en segmentación no era engañoso (lo que frecuentemente ocurre cuando se usan pocos o muchos rangos). A continuación aplicamos una discretización basada en intervalos de igual tamaño que se realizó combinando diferentes intervalos, siendo los más activados (o mayores del total de K_2) los más adecuados para elegir las ROIs tridimensionales (como entrada en *A priori* con *minsup* y *minconf* de 100%). De esta forma, se particionó el rango en un número pequeño de intervalos que tenían una buena coherencia de clase. En particular, como se detalló en la sección 9.2, $K_2 = 5$ se seleccionó como compromiso entre coste computacional y los resultados de exactitud. Como se describió en el capítulo 9, el método propuesto alcanzó ratios de clasificación que mejoraron el caso continuo de partida, en particular *Acc/Sen/Spe* de 96.91/94.64/100 % para SPECT y de 92/84/100 % para FDG-PET. Con-

cluimos que el descubrimiento de las ARs entre ROIs discretas obtenidas de imágenes funcionales puede simplificar el proceso de aprendizaje mejorando la generalización de los resultados aprendidos para el diagnóstico precoz de la AD.

ARs como selector

En cuarto lugar, se presentaron las ARs como una forma efectiva para la selección de características (véase sección 6.5) resolviendo así el problema de la *Maldición de las dimensiones* en SPECT y PET-FDG ADNI para la detección precoz de la AD. La reducción del número de características estuvo condicionada por parámetros como la activación de las ROIs o los umbrales de *minsup* y *minconf* del algoritmo *Apriori* para la extracción de ARs. En primer lugar, tras la obtención de las ROIs (con el mismo tipo de enmascarado y los mismos parámetros que para el caso continuo), las ARs se obtuvieron de los bloques activados de los controles al 100% de *minsup* y *minconf* (para seleccionar los vóxeles más importantes) de la misma forma que en la sección 9.4. La extracción de características y la evaluación se hizo mediante LOO, es decir no se consideró en la extracción de reglas el control que se evaluaba, evitando de esta forma el sobrentrenamiento. Los vóxeles seleccionados se formularon en términos de la activación de los antecedentes y los consecuentes de las ARs combinadas sin repetición. Una de las ventajas es que la dimensión de las características seleccionadas disminuye con ARs de 2445 a 250 para SPECT y de 3776 a 230 para características PET. Se necesitó una reducción a continuación en términos de las técnicas de extracción (PCA y PLS) para establecer los vectores de características principales utilizados como entrada para un clasificador SVM con núcleos lineal para SPECT y rbf para PET. Los puntos de operación convergieron a óptimos de Acc, Spe, Sen para 12 características PLS en SPECT (90.72/89.29/92.68) y PET (90/89.33/90.67), mejorando otras técnicas recientemente publicadas.

Conclusiones conjuntas de los métodos basados en ARs

Destacamos que con todos los métodos presentados en torno a las ARs, hemos conseguido para todas las bases de datos, ratios de exactitud por encima del 90%. En referencia a las ARs entre diferentes tipos de atributos, los mejores resultados los obtuvimos entre atributos discretos. Además conseguimos automatizar el sistema respecto al caso continuo, evitando el uso de umbrales. Por otro lado con las ARs entre atributos discriminantes estudiamos la reducción del coste computacional, asunto de gran importancia en lo que a las

ARs se refiere. El hecho de conseguir que obtuviésemos excelentes resultados en distintas bases de datos con los mismos parámetros, resalta la robustez de los métodos propuestos. Con estos hemos mejorado las ARs entre atributos continuos, el primer paso en la investigación de imagen funcional cerebral mediante ARs, algo que hasta el momento no se había logrado y que ha supuesto un avance dado que se analiza dinámicamente (mediante asociaciones) una alteración neurológica respecto al estado del arte (por ejemplo con VAF-SVM o PCA-SVM) que realiza un análisis estático. En efecto, los ratios de exactitud máximos alcanzados con las comparativas son inferiores a los de los métodos propuestos, por ejemplo (en SPECT y PET respectivamente) de 83% y 81% para VAF y de 86% y de 88.24% para PCA. La introducción de las ARs entre atributos continuos de distintos biomarcadores abre también una nueva línea de estudio para enriquecer los resultados y la caracterización cerebral. Los métodos comentados que asocian mediante ARs distintos tipos de atributos, consiguieron clasificar basándonos en el número de ARs verificadas todos los pacientes normales correctamente, es decir 100% de Spe, dado que las ARs se obtienen del patrón normal porque es más estable y continuo que el patrón AD. Sin embargo, cuando las ARs se emplean como FS, al cambiar a una clasificación mediante kernel SVM, la Spe no alcanza 100% lo que conlleva la consecuente mejora de la sensibilidad (asunto de suma importancia en la detección precoz de una alteración neurológica) dado que los ratios de clasificación siguen superando el 90%. Con la experimentación que en este trabajo se presenta, hemos conseguido demostrar la validez de las ARs como de técnica eficaz e innovadora a día de hoy en el análisis de imágenes funcionales para distintas bases de datos.

11.1.2 LMNN

El algoritmo LMNN realizó una transformación lineal del espacio de entrada mediante la minimización de una función de pérdidas (véase sección 5.4) basándose en la misma observación que la de la regla de decisión kNN ya que clasificaba correctamente un ejemplo si sus k vecinos más próximos compartían la misma etiqueta. En los capítulos 5 y 7, los métodos de aprendizaje de la métrica basada en distancia (LMNN) se investigaron para la clasificación de las imágenes cerebrales SPECT y FDG-PET con la finalidad de mejorar el diagnóstico precoz de la AD. Tras un proceso de enmascarado idéntico al de las ARs para el caso continuo o como selector, como explicó la sección 7.2.1, se obtuvo el valor NMSE de las características tridimensionales cerebrales que fueron seleccionadas posteriormente mediante un t-test. A continuación de la extracción de características, se llevaron a cabo algunos experimentos en el

capítulo 10 para evaluar las ventajas de LMNN como:

- Transformación lineal de los datos reducidos mediante técnicas de extracción de características tales como PLS o PCA, con una clasificación “kernel”-SVM a posteriori.
- Técnica de reducción de características (LMNN-RECT) evaluada con SVM.
- Clasificador mediante distancia euclídea, de Mahalanobis o del método basado en energía.

En los dos primeros casos, con el objetivo de mejorar la exactitud de la clasificación, “kernel”-SVM se utilizó y comparó con los métodos de clasificación basados en LMNN. Además aplicamos el “truco del kernel” para transformar de forma sencilla LMNN en KLMNN con el objetivo de realizar mapeados no lineales, dado que en muchos problemas de aprendizaje de distancias, las distancias euclídeas entre puntos no representaron la topología que intentamos mostrar. El conocido “truco del kernel” (véase sección 7.2.1) puede calcular productos internos entre vectores de características sin construirlos explícitamente [Alipanahi et al., 2008]. KLMNN se consiguió al combinar PCA o PLS kernelizados (mediante “kernel” gaussiano) con LMNN. El primer caso proporciona mejores resultado frente al segundo que contribuye con una reducción del coste computacional del sistema ya que sólo se aplica LMNN para reducir el espacio de características en lugar de una transformación junto con un método de FE. Para el tercer caso, la clasificación LMNN usando modelos basados en energía y distancias Mahalanobis rindió mejor (llegando al 90% de exactitud en SPECT y PET) que cuando se consideraba la distancia euclídea porque ésta sufría un decremento de exactitud conforme el número de características se incrementaba. El método de evaluación utilizado fue k-pliegues que se explicó en la sección de fundamentos teóricos 5.6. LMNN puede verse como el homólogo lógico de SVM con algunas diferencias descritas en la sección 5.5.4. Los métodos propuestos alcanzaron ratios de exactitud para SPECT y PET por encima del 90%, consiguiéndose los mejores resultados ($Acc = 92.78$, $Sen = 91.07$, $Spe = 95.12$ % en SPECT y $Acc = 90.67$, $Sen = 88$, $Spe = 93.33$ % en PET) cuando se utilizaba una técnica de extracción de características basada en la selección NMSE combinada con un t-test, reducción de características PLS y transformación LMNN en combinación con un SVM lineal, mejorando así VAF, PCA o GMM en combinación con SVM. Una de las principales ventajas de nuestras técnicas fue la robustez y estabilidad de los métodos propuestos.

11.2 Trabajo futuro

En esta sección describimos algunas posibilidades de trabajo futuro para continuar mejorando el trabajo iniciado en esta Tesis así como para iniciar nuevas líneas de investigación que lo completen. Las propuestas realizadas se pueden agrupar en los apartados siguientes:

1. **Aplicación en otras enfermedades neurodegenerativas:** Dado que los métodos presentados no dependen de la información patológica de la enfermedad específicamente, serían aplicables a otros tipos de enfermedades neurodegenerativas como por ejemplo:
 - (a) Parkinson, que afecta al 0.3% de la población general y al 3% de la gente sobre los 65 años en todo el mundo [Bradberry et al., 2012].
 - (b) Demencia frontotemporal: es la segunda causa de demencia más común en gente joven (menor de 65 años) y produce atrofia focalizada en los lóbulos frontales y temporales anteriores con una conexión con las características cognitivas [Lillo and Hodges, 2009].
2. **Problemas multiclase:** la enfermedad de AD que nos ha ocupado en este trabajo presenta diferentes estadios según la gravedad de la misma. Dado que la finalidad de esta Tesis era la detección precoz de una alteración neurológica (en particular de la AD), los algoritmos se han aplicado a problemas biclase. El término multiclase se basa en un conjunto de observaciones de varias clases, con el objetivo de aprender una regla de decisión que defina con exactitud nuevas observaciones [Honeine et al., 2013]. Resultaría interesante definir problemas multiclase que permitiesen estructurar con precisión la fase de la enfermedad en la que se encuentra el paciente, así como la diferenciación respecto a otras patologías (en el caso de aplicar el problema a dos o varias anomalías como se describió previamente).
3. **Distintas modalidades de imagen:** En este trabajo se han aplicado los algoritmos a imagen funcional cerebral de tipo SPECT y PET por separado, aunque también podrían aplicarse a otras modalidades de tipo estructural (como MRI) o con dependencia temporal (tales como fMRI) etc. El hecho de fusionar diferentes técnicas de imagen para caracterizar a un paciente permite enriquecer los resultados de clasificación, dado que cada modalidad lleva a cabo un aporte diferente. Es decir, PET se utiliza popularmente en el diagnóstico precoz y en la terapia exacta por su detección no-invasiva, sin embargo, su visualización anatómica no es

tan clara debido a su baja resolución espacial. Dado que este problema puede ser solucionado por técnicas como la tomografía computerizada de rayos X o MRI, la fusión de imágenes resulta el camino más corto para obtener una información triplemente complementaria [qiao Peng et al., 2012]. Otra ventaja de la fusión de imágenes radica en el aspecto económico, dado que existen modalidades de imagen (como PET) que son caras frente a otras más asequibles pero con una menor calidad de imagen (como SPECT). De esta forma, mediante la fusión se conseguiría la misma calidad que en PET pero a menor precio.

4. **Combinación de biomarcadores:** En este trabajo hemos clasificado los biomarcadores para la AD en genéticos, en sangre y/o CSF, así como los biomarcadores de imagen que definen características objetivas extraídas de las imágenes médicas. Sin embargo, la mayoría de los estudios de investigación se centran en una única modalidad de biomarcadores para el diagnóstico de la AD. Aunque en esta Tesis hemos introducido la combinación de biomarcadores de imagen PIB y FDG, resultaría interesante combinar diferentes modalidades de biomarcadores, por ejemplo MRI, FDG-PET y CSF como se describe en [Zhang et al., 2011]. MRI mide la atrofia cerebral, la FDG-PET cuantifica el hipometabolismo y el CSF cuantifica proteínas específicas, por lo que una combinación de biomarcadores aportaría una información complementaria al sistema CAD.

Part VII

Summary in English

Summary

This work is motivated by the study of pattern recognition and classifiers for the early diagnosis of neurologic diseases (in particular such as Alzheimer's Disease, AD) using functional image databases as Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET). With this aim, we propose computer aided diagnosis systems based on two aspects: Association Rules (ARs) and Large Margin Nearest Neighbors (LMNN). On the one hand, ARs facilitate brain study in terms of the regions of interest interconnectivity, in a dynamic and novel way that we have investigated from two points of view: *i*) among different modalities of attributes (continuous, discriminant and discrete) with a classification based on the number of verified rules by each subject and *ii*) as a feature selector with a posterior reduction of the input space (by means of feature extraction techniques) to a kernel-SVM classifier. On the other hand, LMNN algorithm acts as: *i*) a transformer of the input space of features in order to differ between classes under study (normal versus Alzheimer) and to provide a stable classification rate for a wide feature range, *ii*) feature extraction technique by means of a rectangular matrix (LMNN-RECT) and *iii*) a classifier using Euclidean, Mahalanobis distances or energy-based models. With the experimentation, we have demonstrated the effectivity and robustness of the proposed methods (with accuracy rates above 90%) in comparison to other reported ones up to now. In addition, we remark the contribution of different biomarkers, i.e. fluorodeoxyglucose and Pittsburg compound for the early diagnosis of AD.

Chapter 12

Introduction

This chapter describes firstly the objectives and organization of this english summary. Secondly, an introduction to the neurological disease that occupy us in this Thesis (Alzheimer's Disease) is outlined. Finally the obtaining, preprocessing and normalization of the used databases in this work are explained. In particular, we use a SPECT database from the University Hospital Virgen de las Nieves (Granada, Spain), a FDG-PET database from the ADNI Initiative in California (USA), a FDG-PET database from Cartuja Hospital in Sevilla (Spain) and a PIB-FDG biomarker database from ADNI.

12.1 Objectives and organization

The main focus of interest of this Thesis consists in detecting hipoperfusion and hipometabolism patterns for the early diagnosis of neurological diseases, such as the Alzheimer's Disease (AD). For this aim, SPECT (Single Photon Emission Computed Tomography) and PET (Positron Emission Tomography) databases are used. An introduction to AD, the obtaining, preprocessing and normalization of the images are firstly described in this chapter 12. In the following, the two proposed techniques based respectively on Association Rules (ARs) and Large Margin Nearest Neighbors (LMNN) are described. The former (described in chapter 13) presents a baseline in which a novel continuous AR-based classification method is shown [Chaves et al., 2011a]. In addition, some variants of it based on (Fisher Discriminant Ratio) FDR-AR [Chaves et al., 2012c], discretized-AR method [Chaves et al., 2013] or AR as a feature selection method (FS-AR) [Chaves et al., 2012e] are investigated in order to study more specifically associations in the brain and to improve both computational time and/or accuracy rates reaching maxima of 97% (for SPECT) and 92% (for PET).

The latter (described in chapter 14) is based on LMNN algorithm, aimed at the organization of the k -nearest neighbors to the same class, while examples from different classes are separated by a large margin. LMNN is used in this work in different ways: *i*) as a transformation of the feature space obtained by means of PLS (Partial Least Squares) or PCA (Principal Component Analysis) in order to better separate the control subject and AD patient classes, *ii*) as feature reduction technique by performing the transformation as a rectangular matrix (LMNN-RECT), and *iii*) as a classifier comparing Euclidean, Mahalanobis and energy-based methods [Chaves et al., 2012d]. The proposed schemes outperform the current existing methods for the early detection of the Alzheimer's disease. Conclusions and future work are explained in chapter 15.

12.2 Neurodegenerative disorders: AD

Dementia, one of the most severe and frequent neurodegenerative disorders in the elderly population, has important and dramatic health as well as socio-economic implications. Furthermore, the incidence and prevalence of these diseases is increasing due to the aging population, particularly in the United States, Europe, and Japan [Chaves et al., 2012c].

AD, the most common neurodegenerative disease, affects 30 million of individuals worldwide (5.4 million Americans and is believed to cost the economy

385 dollars billion per year). Currently available treatments minimally impact the disease and more effective interventions are needed. Accomplishing this will require a sophisticated understanding of its cause or causes. The definition of AD has varied over the years. Initially, and a number of decades thereafter, AD was considered a form of presenile dementia. Elderly demented individuals were not generally diagnosed with AD, even though their brains frequently contained neuritic plaques and neurofibrillary tangles. Later, the definition was changed to include all patients, regardless of age, who were felt to have a plaque-and-tangle related dementia [Swerdlow, 2011]. Alzheimer's disease (AD) is a chronic, progressive, neurodegenerative disease lasting many years, and unfortunately, placebo-controlled data on AD medications are mainly from short-term studies [Vellas et al., 2007].

AD therapeutics might be divided into symptomatic treatments that ameliorate the symptoms of AD without impacting the underlying disease process and disease-modifying therapies that affect the disease process leading to cell death and slow the clinical course [Cummings, 2009].

12.3 Functional medical images

Neuroimaging technology is advancing at an impressive pace and is having huge fallouts both at the research and at the practical clinical level. At the research level, structural and functional neuroimaging are the unique methodologies allowing the in vivo study of brain pathology at macro and micro-molecular level. Functional brain imaging techniques including SPECT and PET provide functional information, i.e. cerebral sanguineous irrigation or metabolic activity, and enable identifying pathologic anomalies in internal tissues or organs, before anatomical and structural alterations are observable [Villemagne et al., 2005].

SPECT is a noninvasive, 3-D functional imaging modality that can be used to analyze the regional cerebral blood flow (rCBF) in subjects. A SPECT rCBF study is frequently used as a complimentary AD diagnostic tool in addition to the clinical findings. On the other hand, PET measures the rate of glucose metabolism with the tracer [^{18}F] Fluorodeoxyglucose (FDG). In AD, characteristic brain regions show decreased glucose metabolism, specifically bilateral regions in the temporal and parietal lobes, posterior cingulate gyri and precunei, as well as frontal cortex and whole brain in more severely affected patients [Shin et al., 2010].

12.4 Computer aided diagnosis (CAD)

In late-onset AD there are minimal perfusion deficits in the mild stages of the disease, and age-related changes, which are frequently seen in healthy aged people, have to be discriminated from the minimal disease-specific changes. These minimal changes in the images make visual diagnosis a difficult task that requires experienced clinicians.

In order to improve the prediction accuracy especially in the early stage of the disease, when the patient could benefit most from drugs and treatments, CAD tools are desirable. Several approaches for designing CAD systems of the AD can be found in the literature [Fung and Stoeckel, 2007],[Ishii et al., 2006], [Minoshima et al., 1995]. Univariate methodology is based on the analysis of regions of interest (ROIs) by means of some discriminant functions, whereas the second approach (multivariate) is related to statistical analysis techniques. Regarding the first, the most common and used approach is named Statistical Parametric Mapping (SPM) [Friston et al., 2007] software tool and its numerous variants. It was not developed specifically to study a single image, but for comparing groups of images. Regarding multivariate techniques, it is remarkable MANCOVA, which considers as one observation all the voxels in a single scan and requires a higher number of available samples than the one of features. This fact reports the well-known *small sample size* problem that is very common in nuclear medicine studies since the number of images is limited. In this work context, with the clear goal to solve the dimensionality issue, some techniques of feature space reduction were used and combined.

The main goal of CAD systems is to reproduce the knowledge of medical experts in the evaluation of a given subject, i.e. distinguishing AD patients from controls. Thus errors from single observer evaluation are reduced by using a decision support system (DSS) for assisting the identification of early signs of AD. In this sense, several approaches have been proposed for analyzing functional medical images [Chaves et al., 2011a], [Padilla et al., 2010].

12.5 Databases

12.5.1 SPECT database

Baseline SPECT data from 97 participants were collected from the Virgen de las Nieves hospital in Granada (Spain). The patients were injected with a gamma emitting ^{99m}Tc -ECD radiopharmaceutical and the SPECT raw data was acquired by a three head gamma camera Picker Prism 3000. A total of 180 projections were taken with a 2-degree angular resolution. The images of

the brain cross sections were reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter. The SPECT images are first spatially normalized using the SPM software, in order to ensure that voxels in different images refer to the same anatomical positions in the brain allowing us to compare the voxel intensities of different subjects.

In this work, the images have been normalized using a general affine model, with 12 parameters [Friston et al., 2007]. After the affine normalization, the resulting image is registered using a more complex non-rigid spatial transformation model [Salas-Gonzalez et al., 2008]. The deformations are parameterized by a linear combination of the lowest frequency components of the three-dimensional cosine transform bases. A small-deformation approach is used, and regularization is by the bending energy of the displacement field. Then, we normalize the intensities of the SPECT images with respect to the maximum intensity, which is computed for each image individually by averaging over 3% of the highest voxel intensities, similarly as in [Saxena et al., 1998]. After the spatial normalization, one obtains a $95 \times 69 \times 79$ voxel representation of each subject, where each voxel represents a brain volume of $2 \times 2 \times 2 \text{ mm}^3$.

The database is built up of imaging studies of subjects following the protocol of an hospital-based service. First, the neurologist evaluated the cognitive function, and those patients with findings of memory loss or dementia were referred to the nuclear medicine department in the *Virgen de las Nieves* hospital (Granada, Spain), in order to acquire complementary screening information for diagnosis¹. Experienced physicians evaluated the images visually. The images were assessed using 4 different labels: Control (CTRL) for subjects without scintigraphic abnormalities and mild perfusion deficit (AD1), moderate deficit (AD2) and severe deficit (AD3), to distinguish between different levels of presence of hypo-perfusion patterns compatible with AD. In total, the database consists of $N = 97$ subjects: 41 CTRL, 30 AD1, 22 AD2 and 4 AD3 (see table 12.1 for demographic details). Since the patients are not pathologically confirmed, the subject's labels possess some degree of uncertainty, as the pattern of hypo-perfusion may not reflect the underlying pathology of AD, nor the different classification of scans necessarily reflect the severity of the patients symptoms. However, when pathological information is available, visual assessments by experts have been shown to be very sensitive and specific labeling methods, in contrast to neuropsychological tests [Jobst et al., 1998]. Given that this is an inherent limitation of 'in vivo' studies, our working-assumption

¹Clinical information is unfortunately not available for privacy reasons, but only demographic information

Demographic Details			
	Num. of Samples	Sex (M/F) (%)	Age μ [range/ σ]
CTRL	41	32.95/12.19	71.51 [46-85/7.99]
AD 1	30	10.97/18.29	65.20 [23-81/13.36]
AD 2	22	13.41/9.76	65.73 [46-86/8.25]
AD 3	4	0/2.43	76[69-83/9.90]

Table 12.1: Demographic details of the SPECT dataset. AD 1 = mild perfusion deficit, AD 2 = moderate deficit, AD 3 = severe deficit. μ and σ stands for population mean and standard deviation respectively.

is that the labels are true, considering the subject label positive when belonging to any of the AD classes, and negative otherwise. This work does not imply any experimental intervention and has been performed under the approval and supervision of the Clinical and Investigation Ethical Commission of the University Hospital Virgen de las Nieves (CEIC).

12.5.2 FDG-PET database

PET data was obtained from the ADNI Laboratory on NeuroImaging (LONI, University of California, Los Angeles) (<http://www.loni.ucla.edu/ADNI/>). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 60 million, 5-year public- private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co- investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be

followed for 2 years. For up-to-date information, see www.adni-info.org.

FDG PET scans were acquired according to a standardized protocol. A 30-min dynamic emission scan, consisting of 6 5-min frames, was acquired starting 30 min after the intravenous injection of 5.0 ± 0.5 mCi of ^{18}F -FDG, as the subjects, who were instructed to fast for at least 4 h prior to the scan, lay quietly in a dimly lit room with their eyes open and minimal sensory stimulation. Data were corrected for radiation-attenuation and scatter using transmission scans from Ge-68 rotating rod sources and reconstructed using measured-attenuation correction and image reconstruction algorithms specified for each scanner. Following the scan, each image was reviewed for possible artifacts at the University of Michigan and all raw and processed study data was archived. Subsequently, the images were normalized through a general affine model, with 12 parameters using the SPM5 software [Friston et al., 2007]. After the affine normalization, the resulting image was registered using a more complex non-rigid spatial transformation model. The non-linear deformations to the Montreal Neurological Imaging (MNI) Template were parameterized by a linear combination of the lowest-frequency components of the three-dimensional cosine transform bases [Ashburner and Friston, 1999]. A small-deformation approach was used, and regularization was by the bending energy of the displacement field, ensuring that the voxels in different FDG-PET images refer to the same anatomical positions in the brains. After spatial normalization, an intensity normalization was required in order to perform direct images comparisons between different subjects. The intensity of the images was normalized to a value I_{max} , obtained averaging the 0.1% of the highest voxel intensities exceeding a threshold. The threshold was fixed to the 10th bin intensity value of a 50-bins intensity histogram, for discarding most low intensity records from outside-brain regions, and preventing image saturation. Participant's enrolment was conditioned to some eligibility criteria. General inclusion-exclusion criteria were as follows:

- Normal control subjects: Mini Mental State Examination (MMSE) scores between 24–30 (inclusive), a Clinical Dementia Ratio (CDR) of 0, non depressed, non MCI, and non demented. The age range of normal subjects will be roughly matched to that of MCI and AD subjects. Therefore, there should be minimal enrolment of normals under the age of 70.
- Mild AD: MMSE scores between 20–26 (inclusive), CDR of 0.5 or 1.0, and meets NINCDS/ADRDA criteria for probable AD.

The PET database collected from ADNI consists of 150 labeled PET im-

Demographic Details			
	Num. of Samples	Sex (M/F) (%)	Age μ [range/ σ]
CTRL	75	58.67/41.33	75.97 [62-86/4.91]
AD	75	62.67/37.33	75.72 [55-88/7.40]

Table 12.2: Demographic details of the PET dataset. μ and σ stands for population mean and standard deviation respectively.

ages: 75 control subjects and 75 AD patients (see table 12.2 for demographic details). ADNI patient diagnostics are not pathologically confirmed, introducing some uncertainty on the subject's labels. Using these labels, allows to test the robustness of the classifier. This should be also considered when comparing to other methods tested on autopsy confirmed AD patients, on which every classifier is expected to improve its performance [Illán et al., 2011a]. Written informed consent was obtained from all ADNI participants before protocol-specific procedures were performed. The informed consent not only covers consent for the trial itself, but for the genetic research, biomarker studies, biological sample storage and imaging scans as well. The consent for storage includes consent to access stored data, biological samples, and imaging data for secondary analyses. By signing the consent, ADNI participants authorize the use of the data for large scale, multicenter studies that combine data from similar populations.

12.5.3 PIB-PET database

PET data was obtained from the ADNI Laboratory on NeuroImaging (LONI, University of California), website (<http://www.loni.ucla.edu/ADNI/>). Determination of sensitive and specific markers of early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The used database consists of baseline [^{18}F]-FDG and PiB PET data from the same 103 ADNI participants conditioned to inclusion/exclusion criteria which were acquired from different scanners. PET data was separated into 3 different classes: 19 CTRL, 65 MCI and 19 AD.

Preprocessing

Every subject PET image is co-registered to the AC-PC line, averaged, standardized spatially and in intensity, and with uniform resolution in order to make possible a direct comparison between images [Illán et al., 2010]. Sub-

jects were spatially normalized using SPM5 software normalizing each [^{18}F]-FDG image volume with a general affine model with 12 parameters and a cost function to the MNI template. One of these parameters were obtained for the [^{18}F]-FDG image. The same preprocessing for both modalities is guaranteed as each coregistered PiB image was normalized using them.

12.5.4 PET Cartuja database

The PET database consists of 60 images (18 normal and 42 severe AD subjects) provided by the PET Cartuja Clinic (Seville, Spain). The labels of the images were assigned by a group of experienced clinicians and are assumed to be the *gold standard* since it is impossible to know for sure if a subject is suffering the disease until a post-mortem analysis is done [Segovia et al., 2010]. The radiopharmaceutical used was 18F-FDG and the camera was a Siemens CTI ECAT EXACT 47. The post-filter step for PET images consists of a Gaussian filter and an axial filter. To ensure that the same position in the volume coordinate system in different images corresponds to the same anatomical position, all images have been spatially normalized using SPM yielding a $69 \times 95 \times 79$ 3D functional activity map for each patient. Moreover, the intensity of voxels was normalized by equalizing the maximum value and scaling the remaining ones.

Chapter 13

Association Rules

This chapter presents a novel AR-based method from various points of view. Firstly, ARs are investigated among different kind of attributes: continuous, discriminant based on Fisher Discriminant Ratio (to reduce the computational cost) and discrete (to avoid the use of thresholds in the ROI extraction and to make the CAD system more automatic). The classification is performed taking into account the number of verified rules. FDG and PIB biomarker combination is also studied in order to enrich the classification. In addition, ARs are introduced as a feature selection method (FS-AR) followed by a feature reduction of the space used as input to a kernel SVM classifier. Different experiments are shown for databases SPECT, FDG-PET (ADNI), PET-Cartuja and PIB-FDG PET.

13.1 Introduction

Data mining lets the discovering of novel, useful and hidden knowledge from databases [Menzies and Hu, 2003]. Association Rule (AR) is a powerful area of data mining with a wide range of applications in many areas of business practice and also in research. In particular, ARs have attracted the attention of practitioners with applications such as customer relation management, market basket, economic and financial-time series analysis, etc. For example, customers who order product A often also order product B or C or employees who said positive things about initiative X also frequently complain about issue Y but are happy with issue Z [Srikant and Agrawal, 1995].

ARs, that were firstly proposed in [Agrawal et al., 1993], enable to find the associations and/or relationships among inputs items of any system in large databases and later eliminate some unnecessary inputs. Medical image mining is a promising area of computational intelligence applied to automatically analyze patient's records aiming at the discovery of new knowledge potentially useful for medical decision making [Ordoñez et al., 2000]. Some approaches using ARs have been proposed in [Antonie et al., 2001], [Zaiane et al., 2002] for tumour classification in mammograms. On the other hand, other techniques explore AR-mining for the early diagnosis in lung cancer [Zubi and Saad, 2011]. However, the application of ARs to neuroimaging is still a challenge, for example in Alzheimer's Disease. The key issue for these applications is that available image databases provide a great environment to apply some image recognition methods to extract the useful knowledge, i.e. rules to characterize the normal and abnormal patterns of the diseases in a medical decision making scenario as in [Chaves et al., 2011a].

13.2 ARs among continuous attributes

This section describes AR method among continuous attributes. First of all, *Apriori* algorithm is described for the AR-mining that uses as input a set of continuous attributes or ROIs obtained with the combination of VAF and Activation Estimation techniques. Finally classification is performed taking into account the number of verified ARs by each subject. In addition, biomarker combination with ARs is introduced as an interesting way to enrich the results.

13.2.1 AR-mining

Algorithms for the AR-mining such as AIS, SETM, Apriori etc [Agrawal et al., 1993], [Houtsma, 1995], [Borgelt, 2005], [Wang et al., 2005], [Kuo et al., 2011]

make multiple passes over the data. In this work, we apply Apriori [Srikant and Agrawal, 1995], which is the most popular one and a state-of-the-art algorithm since most of the AR algorithms are variations of this. In the first pass, the support of individual items is counted determining which of them are large, i.e. have minimum support (*minsup*). In each subsequent pass, we start with a seed of large itemsets found to be large in the previous pass and we use them for generating new potentially large itemsets called candidate itemsets, and count the actual support for these candidate itemsets during the pass over the data. At the end of the iteration, the candidate itemsets which are actually large are determined to become the seed for the next pass and so on until no new large itemsets are found.

In our context, let m be the number of activated regions (also called ROIs) $I = i_1, i_2, \dots, i_m$ and let T be a database of transactions $T = t_1, t_2, \dots, t_n$ that relate only the ROIs of controls. Control patterns are less variable than AD patterns [Illán et al., 2010], thus they are more suitable for obtaining relevant relationships among brain regions using ARs, which will not be satisfied in general by AD patients. Each transaction t is represented as $X \Rightarrow Y$, where $X, Y \subset ROIs$ of controls and $X \cap Y = \emptyset$. In our problem, an AR: $X \Rightarrow Y$, which relates two activated items (regions) of the normal pattern, is defined as “activation in area X implies activation in area Y ” given some certain support and confidence constraints. The rule $X \Rightarrow Y$ has a support s if $s\%$ of the transactions in T contain $X \cup Y$ and a confidence c , if $c\%$ of the transactions in T that contain X also contain Y , [Wu et al., 2004] that is:

$$confidence(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X)} \quad (13.1)$$

For example, if the rule $X \Rightarrow Y$ has a confidence of 1 in the database means that for 100% of the transactions containing both ROIs, the rule is correct. While confidence is a measure of the rule’s strength, support corresponds to statistical significance. Support and confidence do not measure association relationship but how-often patterns co-occur and only make sense in the AR-mining.

13.2.2 Input for AR-mining algorithm

In order to obtain the activated ROIs (which are the inputs for Apriori), we propose to apply a combination of Voxels as Features (VAF) and threshold-based Activation Estimation (AE) methods for feature extraction of each subject. The former method (VAF) is a way of including all the voxel intensities inside the brain as features, thus no explicit knowledge about the disease is

needed and the inclusion of *a priori* information about the pathology into the system is avoided. Hence this information is implicitly obtained from image databases [Stoeckel et al., 2004]. The AE method leads to a reduced list of activation maxima containing only those maxima which have one or more other maximums in their vicinity (activated voxels). In functional imaging, each voxel carries a grey level intensity $I(\mathbf{x}_j)$, which is related to the regional cerebral blood flow [Holman et al., 1992], glucose metabolism [Ng et al., 2007], etc. in the brain of a patient, depending on the image acquisition modality.

Voxel selection by means of AR mining is addressed by first dividing each image into 3D $v \times v \times v$ cubic blocks. The number of blocks in which an image is divided depends on the size of the blocks (v) and the overlap between adjacent blocks, g , defined in terms of number of voxels. In this way, the centers of the blocks considered define a 3D $g \times g \times g$ grid, considering the possibility of some overlap between adjacent blocks. AE parameterizes a block by the proportion of activated voxels t .

13.2.3 AR-based classifier

The stages of the AR-based technique [Chaves et al., 2011a] are:

- Training stage: ROIs were previously extracted for each subject in the database as described in section 13.2.2. In the training stage, ARs are mined only from controls using cross-validation strategy [Gidskehaug et al., 2008], [Stone, 1974]. Whenever a specific control is being evaluated in the test stage, it is not considered in the AR-mining, avoiding the overtraining. If the availability of samples is limited, separation of the data into a training and validation set may decrease the quality of both the calibration model and the validation. Therefore, this is solved by processing all available information (i.e. scans), in order to estimate the residual error of the model in the test stage [Gidskehaug et al., 2008] in an iterative leave-one-out (loo) process.
- Test stage: In this stage we check the number of ARs which are verified by each test subject following the loo validation strategy. If a subject verifies more or an equal number of ARs than a *class-threshold* then, it is classified as NOR otherwise as AD.

13.2.4 Biomarker combination and ARs

Traditionally, changes in glucose metabolism in the human brain have been measured by PET imaging using [18F] FluoroDeoxyGlucose (FDG), which

has long been used to visualize hypometabolism in patients with AD. In addition, recently developed radiotracers such as (11)C-Pittsburgh Compound B (PIB) has made it possible to visualize amyloid deposits, and potentially neurofibrillary tangles in vivo in human subjects [Shin et al., 2010].

ARs are proposed in this work as a methodology which takes the associations and/or relationships in PET data considering the possibility of a diagnostic approach predicated on the concept of certain combinations of biomarkers providing complimentary information [Weiner et al., 2012]. ROIs of the same subjects (FDG and PIB) are merged without repetition in order to study the effect of the biomarker (FDG+PIB) combination. The ROIs obtained are used as input in the AR-mining stage (from CTRL), which is performed by means of an Apriori algorithm that uses different *minconf* and *minsup* values as input.

Voxel-based analysis of PIB-PET shows that subjects with AD compared with age matched controls show highly significant PIB retention in the frontal, parietal, temporal, and posterior cingulate/precuneus cortices. Statistically significant glucose hypometabolism in the FDG-PET analysis occurred in similar brain regions with highly significant PIB retention in the PIB-PET analysis.

13.3 ARs over discriminant regions

Aiming to reduce the computational cost of the system described in section 13.2, we have applied Fisher discriminant ratio in order to obtain the most discriminant regions to be related with ARs.

13.3.1 Fisher discriminant ratio

Discriminant analysis [Fisher, 1936] addresses the following question: Given a data set with two classes, say, which is the best feature or feature set (either linear or non-linear) to discriminate the two classes? [Mika et al., 1999]. The Fisher Discriminant Ratio (FDR) criterion is characterized by its separation ability as shown in [Webb, 2002]. For the two-class case, it may be defined as follows:

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (13.2)$$

where μ_i and σ_i^2 denote the i -th class mean value and variance for each input variable (voxel), respectively. For a given voxel, the ratio value grows as the difference of the mean values of each two classes increases or the cumulative

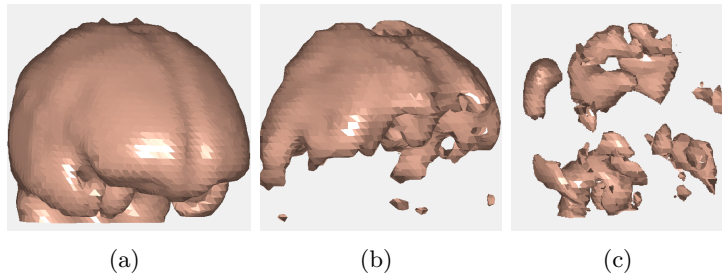


Figure 13.1: 3D Mask: a) average of controls for SPECT database, b) FDR-mask for SPECT database, and c) FDR-mask for PET database.

scattering in each class decreases, thus being useful to reveal discriminant voxels.

13.3.2 FDR-based masking

A FDR-based masking process is performed in this work in order to identify discriminant voxels for AR mining [Chaves et al., 2012c]. This step not only improves the accuracy of the proposed method but also reduces the computational cost as the selection of voxels that exceeds the FDR threshold lets reduce the dimensionality of the problem, which means a lower number of input variables to the AR mining process. In the case of the functional images, the voxels that satisfy a particular FDR threshold level are selected as the most discriminant variables [Illán et al., 2011b]. Figure 13.1 shows the masks for a) the average of controls as a reference, b) FDR-based SPECT mask and c) FDR-based PET mask. The latter images were obtained at a FDR threshold named h equal to 0.2, that is, just the voxels with an FDR value above 0.2 of the maximum are considered for further analysis. Note that figure 13.1.a shows all the voxels, which are inside the skull, to be considered for the AR mining whereas in figures 13.1.b and 13.1.c the specific regions, depicted by the FDR criterion and used for AR discovery afterwards, are shown. This dimension reduction provides good classification results together with lower computational cost in the light of experimental results (see section 13.6.2).

13.4 ARs among discrete attributes

This section set the basis of discretization techniques in order to introduce the idea of ARs among discrete attributes. Properly addressing the discretization process is an important problem as most real-world application of classification

learning involve continuous-valued attributes [Fayyad and Irani, 1993]. The attributes in a learning problem may be nominal (categorical) or continuous (numerical). The term continuous is used in the literature to refer to attributes taking on numerical values (integer or real); or in general an attribute with a linearly ordered range of values.

The process of partitioning continuous variables into categories is usually termed discretization whose first goal is to find a set of cut points to partition the range into a small number of intervals that have good class coherence. In addition, feature discretization is closely related to dimensionality reduction [Chaves et al., 2013]. The second goal is to minimize the number of intervals without significant loss of attribute mutual independence [S. Kotsiantis, 2006]. In [Liu et al., 1998a], it is proposed the discretizing of continuous attributes based on the classification predetermined class target in order to integrate classification and AR mining. Generally, the discretization methods can be categorized as: supervised or unsupervised, direct or incremental and static or dynamic [Kohavi, 1995]. While local methods produce partitions that are applied to localized regions of the instance space; global methods, such as binning, produce a mesh over the entire n -dimensional continuous space [Chmielewski and Grzymala-Busse, 1994]. In contrast to unsupervised, supervised methods make use of the class labels in the discretization process. Whereas static methods, such as binning or entropy-based partitioning [Fayyad and Irani, 1993], perform one discretization pass of the data of each feature and determine the value of the maximum number of intervals (K); dynamic methods conduct a search for all the features simultaneously in order to capture interdependencies in feature discretization. Equal width interval binning is an unsupervised method often applied as a means for producing nominal values from continuous ones and involves dividing the range of observed values of a continuous feature into K (selected by the user) equally sized bins [Dougherty et al., 1995]. If a variable x is observed to have values bounded by x_{min} and x_{max} , then this method computes the bin width as:

$$\alpha = \frac{x_{max} - x_{min}}{K} \quad (13.3)$$

and constructs bin boundaries at $x_{min} + i \cdot \alpha$ where $i=1, \dots, K-1$.

Our adaptation involves discretizing continuous attributes (mean 3D voxel intensity) into histogram bins and is also motivated by an equal width interval binning [Dougherty et al., 1995]. Combined or independently considered discretized bins define the value that the feature must possess to be considered in the AR-mining stage. Properly discretized data can simplify the process of learning and may improve the generalizability of the learned results.

13.5 AR-based feature selector

Feature selection and extraction, which are based on a limited training set, are likely to display poor generalization performance on new datasets. To address this challenge, a novel voxel selection method based on association rule (AR) mining is proposed for designing a computer aided diagnosis (CAD) system. This section describes the benefits of feature selection applied to ARs, secondly the feature extraction and classification are described.

13.5.1 Feature Selection

The high dimensionality of the input feature space in comparison with the relatively small number of subjects [Costafreda et al., 2009] is a widely known problem called *curse of dimensionality*. Many of these features are irrelevant or redundant. Unnecessary features increase the size of the search space and make generalization more difficult. This curse of dimensionality (each feature is a separate dimension) is a major obstacle in machine learning and data mining, so some form of feature selection is often applied in order to solve it [Gheyas and Smith, 2010], [Salas-Gonzalez et al., 2010a].

In addition, FS has benefits as it speeds up the testing process and increases classification accuracy. The development of feature selection has two major directions: subset selection and feature ranking. In the feature subset selection problem, a learning algorithm is faced with the problem of selecting a relevant subset of features upon which to focus its attention, while ignoring the rest [Kohavi, 1995]. Subset selection typically falls into three categories: filters, wrappers and embedded [Maldonado et al., 2011]. Filter method uses statistical properties of the features in order to filter out poorly informative ones. Wrapper explores the whole set of variables to score feature subsets according to their predictive power, optimizing a performance criterion of the subsequent algorithm that uses the respective subset for classification [Kohavi and John, 1997]. The last approach (embedded methods) performs FS in the process of model construction. For example, an embedded method for feature selection using SVMs is proposed in [Maldonado et al., 2011]. As feature ranking methodology is concerned, the method makes use of ranking information, instead of simply viewing the ranks as flat categories. The final objective is to find a set of features with maximum importance and minimum similarity, that is, it tries to avoid selecting redundant features providing a greedy search algorithm to solve the optimization problem [Geng et al., 2007].

13.5.2 ARs for feature selection

FS is aimed to select the hopefully best subset of features of the input feature set. As an example, ARs were proposed in [Karabatak and Ince, 2009] as an FS methodology which takes the associations and/or relationships in erythematosquamous disease data eliminating in this way unnecessary inputs. Feature are obtained from controls following first the same procedure as in 13.2.2 for ROI extraction as input for Apriori and in the following AR-mining using as input *minsup* and *minconf* parameters. The voxel selection algorithm considers just the voxels that are involved as antecedents or consequents of the ARs mined in the previous step. Voxels are selected by AR combination stage, that is, the antecedents and consequents of previously obtained ARs are merged without repetition and the output of the AR-enabled voxel selection method is the activation level of these ROIs [Chaves et al., 2012e].

Feature extraction and classification

After the AR-based FS process, it is necessary to reduce more the feature space used as input to the classifier. Feature extraction (FE) creates new features based on transformation (linear or nonlinear) from the original features to a lower dimensional space. Dimensionality reduction is performed by means of Principal Component Analysis (PCA) [Jolliffe, 2002] or Partial Least Squares (PLS) [Gerlach et al., 1979]. The resulting set of features may provide a better discriminant ability than the subset of given features selected by ARs. PCA generates an orthonormal basis vector that maximizes the scatter of all the projected samples, which is equivalent to finding the eigenvalues of the covariance matrix [Pérez, 2010]. PLS represents a wide class of methods for modeling relations between sets of observed variables by means of latent variables [F.Segovia, 2010]. Classification is performed by a SVM classifier [Bishop, 2006] which has been successfully used in a number of applications [Ramírez et al., 2009b], [Chaves et al., 2009b]. Moreover, recent developments in defining and training statistical classifiers make it possible to build reliable classifiers in very small sample size problems and even may find non-linear decision boundaries for small training sets.

13.6 Experiments

13.6.1 Experiments of AR among continuous attributes

Several experiments were conducted in order to evaluate the combination of VAF and AE feature extraction methods aiming to posterior AR mining for

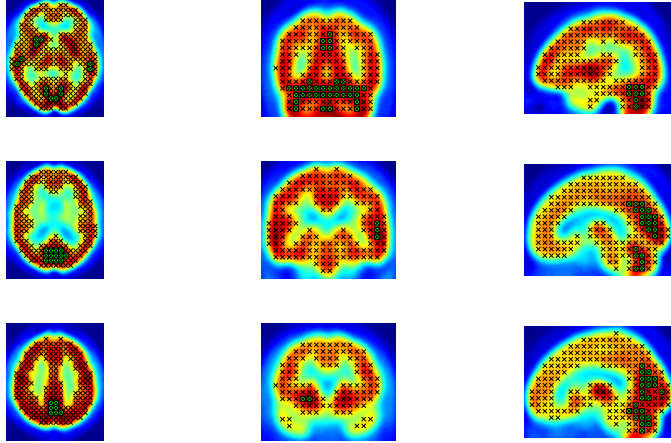


Figure 13.2: Bidimensional activated regions (rectangles) from axial, sagittal and coronal axis with $minsup$ and $minconf$ of 100%, $t=90\%$. SPECT database, AR-continuous-based method.

brain image classification. In the first step (VAF) [Stoeckel et al., 2004], a mean tridimensional image is computed by averaging all control images from the SPECT database. An activation threshold intensity a_T is fixed by the 50% of the maximum intensity of this mean image, defining a 3D mask. Voxels outside this mask are discarded from the database images, and therefore voxels outside the brain and poorly activated regions are excluded from this analysis. Then, an image voxel is considered as *activated* when its intensity value I is above the activation threshold, that is, when $I > a_T$.

In all the experiments, 3D activated blocks (centered in mask coordinates) of size $9 \times 9 \times 9$ are selected, which include voxels above an intensity threshold as described in section 13.2.2. Coordinates of the blocks are sampled by means of a $4 \times 4 \times 4$ 3D grid to reduce the computational complexity of the method. It was experimentally proven that a region is considered of interest or *nearly-activated*, if more than 90% or a fraction (650 voxels) of the total 729 voxels are activated. Figure 13.2 shows transaxial, coronal and sagittal slices of a mean SPECT image illustrating the $4 \times 4 \times 4$ grid locations (crosses) and the activated regions involved in ARs (squares) in terms of antecedents and consequents. ARs are mined using a high $minsup$ and $minconf$ (100%) to extract the rules that better distinguish AD subjects from controls. Note that

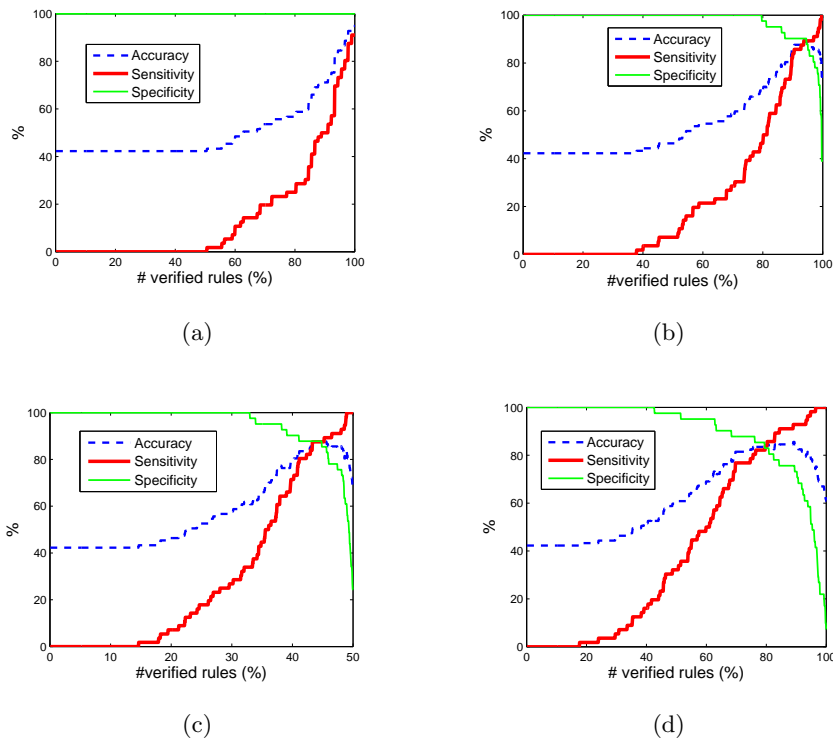


Figure 13.3: AR-continuous-based method. Accuracy, Sensitivity and Specificity versus the percentage of verified rules when the activation block-threshold is 90% for *minsup* and *minconf* of a) 100%, b) 95%, c) 90% and d) 80%. Note that the overall number of mined rules and the Maximum accuracy rates obtained are a) 29756 rules, 94.87%, b) 84396 rules, 89.69%, c) 146768 rules, 87.63% and d) 288190 rules, 85.57%

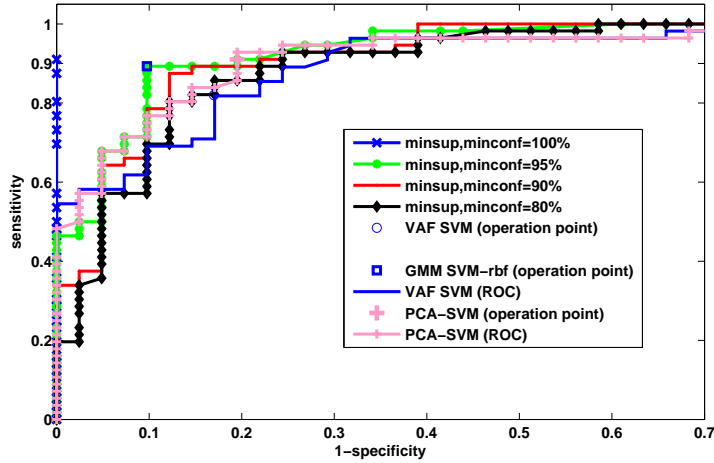


Figure 13.4: ROC Analysis (SPECT). AR-continuous-based method: AR method with different selected percentages of *minsup* (100%, 95%, 90%, 80%) and *minconf* (100%, 95%, 90%, 80%) in the rule extraction. Comparison to other recently reported methods represented by their operation points. Observe the peak value of the Acc =94.87%, with Sen= 91.07% and Spe= 100% (cross-shaped mark), for *minsup*=*minconf*=100. The AUC obtained for each ROC is: ARs method (*minsup*: AUC) 100%: 0.9554, 95%: 0.9371, 90%: 0.9233, 80%: 0.9027; VAF SVM: 0.8993; PCA SVM: 0.9177

temporoparietal brain regions, which are characteristic brain regions for AD, are considered as activated ROIs by the algorithm. If a higher activation block threshold is considered, the results are quite good although important ROIs may be discarded. By contrast, if a lower activation block-threshold is selected, the Accuracy results are lower with an increase in computational time. The performance of the AR-based classifier, as a tool for the early detection of AD, was evaluated in depth in terms of Accuracy, Sensitivity and Specificity. The stages of this AR-continuous-based system are summarized in section 13.2.3.

Mined ARs are discriminant for the early diagnosis of AD when the high-est *minsup* and *minconf*, i.e. 100%, are selected, otherwise we include hundreds of data-dependent rules which do not effectively model the normal pattern. Moreover, the decrease of Specificity at a determined n_0 value in figures (13.3(b) with *minsup*, *minconf* of 95% at n_0 =95% of verified rules, 13.3(c) with *minsup* and *minconf* of 90% at n_0 =88% and in 13.3(d) with *minsup*, *minconf* of 80% at n_0 =79.5%) corroborates this assumption. If *minsup* and

minconf values are lower than 100%, the method provides an increasing detection ability up to an specific n_0 value. Then, when a $n > n_0$ value is considered Specificity decreases noticeably. In addition, this may lead to an increase in the computational cost of the CAD system along with a penalty in the maximum accuracy rate (89.69% in figure 13.3(b), 87.63% in figure 13.3(c) and 85.57% in figure 13.3(d)) because non-discriminant rules are also included in the classification process. In the light of figure 13.3, we notice that extracting rules at the highest *minsup* and *minconf* percentage, yields the best results in terms of Accuracy, Specificity and Sensitivity (94.87%, 100% and 91.07% respectively).

The ROC curve is aimed to compare the compromise that can be made between Sensitivity (TPF, True Positive Fraction) and 1-Specificity (FPF, False Positive Fraction) for providing a complete description of disease detectability [Metz, 1978]. Figure 13.4 shows the ROC curves for different *minsup* and *minconf* values (from 100% to 80%) which evaluate the trade-off between sensitivity and specificity as the threshold of ARs verified by each subject is increased. It has been established a comparison to other recently reported CAD methods such as the ones based on VAF [Stoeckel et al., 2004] (ROC curve and operation point), Principal Component Analysis (PCA) [Pérez, 2010] (ROC curve and operation point) and Gaussian Mixture Model (GMM) [F.Segovia, 2010] combined with Support Vector Machines (SVM), which is represented by its operation point. The ROC curve of the proposed method is shifted up and to the left outperforming the baseline used in this paper as shown in figure 13.4. Due to the fact that Specificity is always 1, avoiding to misclassify any control, when the maximum *minsup* and *minconf* (100%) are used, the ROC curve in this case reaches 0.91 Sensitivity approaching the ideal CAD system (1 sensitivity, that would mean that every control and AD would be correctly classified). In this set up, our AR-based CAD system correctly classified all controls (41) and all AD patients but 5 scans (4 labeled AD1 and 1 labeled AD2) as expected, since AD1 pattern is still a challenge to be diagnosed. If we only consider the case NOR vs AD1 the precision rates of the method are Acc=94.37%, Sen=86.67%, Spe=100% which still represents a great advance in the field.

If a subject verifies more or an equal number of ARs than a class-threshold, it is classified as control otherwise as AD. This means that as the threshold of verified rules is increased, a smaller number of cases would be classified as control, i.e., the Specificity should decrease. By contrast, this does not happen when the highest *minsup* and *minconf* (100%) are selected in the AR-mining, that is, when the rules perfectly model the normal pattern as being

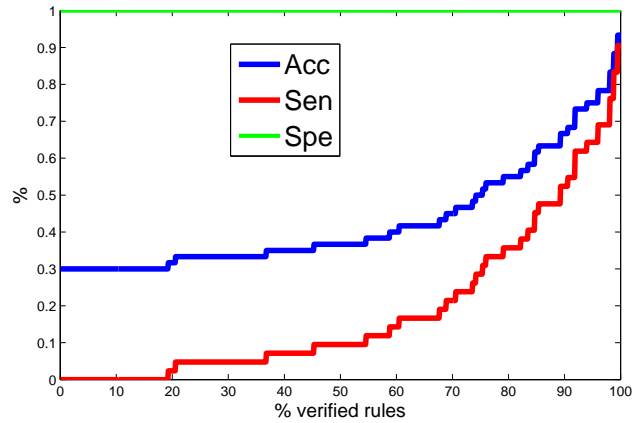
<i>minconf</i> , <i>minsup</i>	100%	95%	80%
% verified ARs	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)
70	53.71 / 19.64 / 100	58.76 / 28.57 / 100	81.44 / 76.79 / 87.80
90	71.13 / 50 / 100	86.60 / 83.93 / 100	84.54 / 92.86 / 73.17
100	94.85 / 91.07 / 100	74.23 / 100 / 39.02	60.82 / 100 / 0.07
	VAF	PCA	GMM
	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)	Acc / Sen / Spe (%)
	83.51 / 83.93 / 82.93	86.56 / 91.07 / 80.49	89.69 / 90.24 / 89.29

Table 13.1: AR-continuous-based method. Statistical measures of performance of AR-based feature selection methods for the sample group using several set-ups (*minconf*, *minsup* and % verified rules). Voxel-as-features (VAF), GMM and PCA operation results are reported as reference. VAF parameters: linear SVM classifier, GMM parameters: $\sigma=6$ RBF-SVM classifier with 8 components and PCA parameters: $\sigma=6$ RBF-SVM classifier with 16 components.

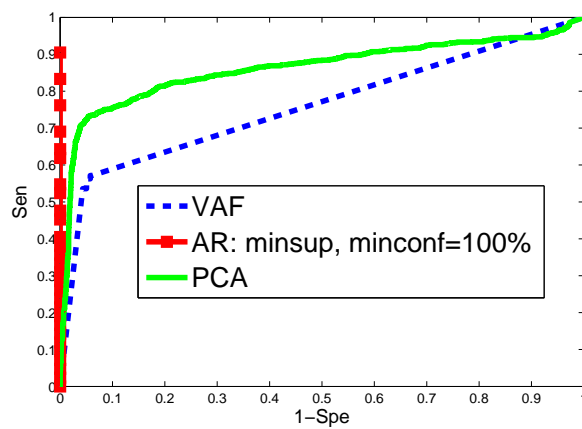
the most discriminative for AD. The best classification results are obtained when 100% of the rules are verified. These results are in agreement with the expected behavior of the system since normal subjects are assumed to have a common SPECT image pattern and to verify most of the ARs mined. This fact supports the idea that it is preferable to only use ARs mining in a set of controls. In table 13.1 we summarize the statistical measures obtained with the proposed methods and compared them to recently reported CAD systems [López et al., 2009a], [Segovia et al., 2010].

As a conclusion, the analysis of the ROC curves shows that the proposed CAD system based on AR mining yields the best trade-off between sensitivity and specificity by shifting the operating point up and to the left in the ROC space [Metz, 1978]. As expected, best results are obtained with the proposed AR technique for the highest *minsup* and *minconf* (of 100%).

For the analysis of PET Cartuja database, ROIs with the same size (v, g) and with the same activation (t) than in SPECT have been used. Figure 13.5.a shows Acc, Sen and Spe versus the percentage of verified rules. As *minsup* and *minconf* are of 100%, the maximum rates (Acc=93.33%, Spe=100%, Sen=90.48%) are reached at 100% of verified rules, that is, 82082. Figure 13.5.b shows ROC curve of the proposed method that outperforms PCA and VAF combined with SVM. These results are in agreement with the AUC described in caption of the figure 13.5.



(a)



(b)

Figure 13.5: AR-continuous-based method, PET- Cartuja database. a) Acc, Spe y Sen versus the percentage of verified rules (total AR number=82082). b) ROC curve in comparison with other reported methods. AUC: VAF(0.7592), PCA(0.8683), AR(0.9158).

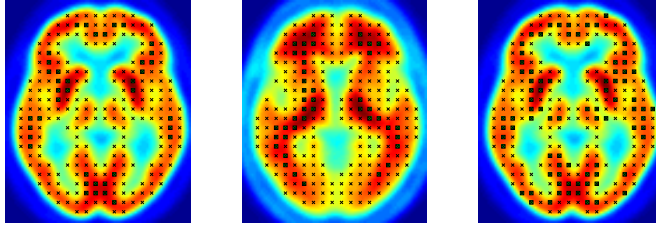


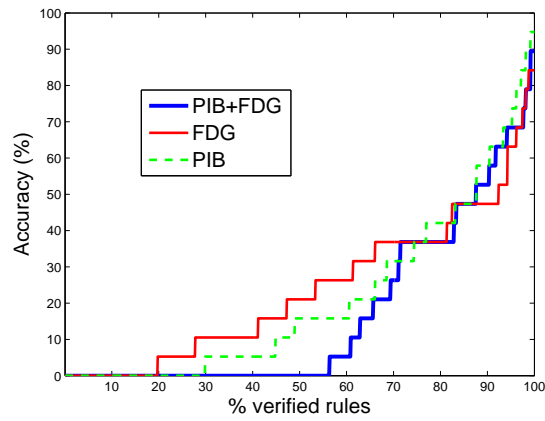
Figure 13.6: ARs depicted (as squares) in coronal slices at minsup , $\text{minconf}=100\%$ for PET database: a) FDG, b) PIB, c) FDG plus PIB (over FDG mean image) for the CTRL versus AD set. AR-continuous-based method, biomarker database.

Experiments of biomarker-AR-based system

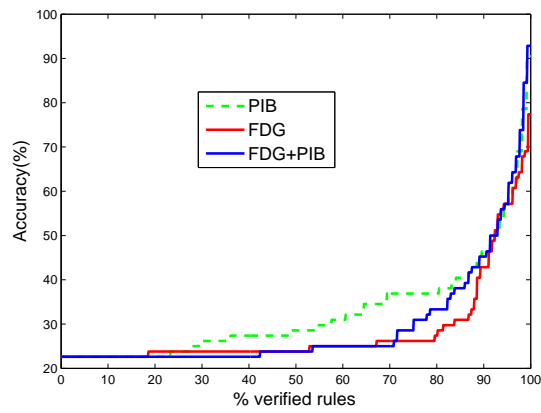
In order to remark the robustness of the method, we use the same parameters (v, g, t) than for the SPECT and PET- Cartuja databases. The experiments considered training and testing the system in two different sets: AD versus CTRL (set 1) and CTRL versus MCI (set 2). The CAD system is analyzed using separately $[^{18}F]$ -FDG or PIB biomarker and the combination of both. In the latter case, the ROIs of the same subjects (FDG and PIB) are merged without repetition in order to study the effect of the biomarker ($FDG + PIB$) combination. The ROIs obtained are used as input in the AR-mining stage, which is performed by means of an Apriori algorithm that uses as input minconf and minsup values [Chaves et al., 2011a]. AR-mining and classification are described in sections 13.2.1 and 13.2.3.

Figure 13.6 shows for the CTRL versus AD set, the coronal slices of a mean control PET image (or mask), being FDG-PET mask depicted in figures 13.6.a and 13.6.c, and PIB-PET mask in figure 13.6.b illustrating the $4 \times 4 \times 4$ grid locations (crosses) and the activated regions involved in ARs (squares) in terms of antecedents and consequents of the rules. Note that, when combined both FDG and PIB biomarkers on FDG mask (figure 13.6.c), more crosses appear depicted as some new rules were mined outperforming results with just FDG ROIs.

If 100% of minsup and minconf are used, maxima results are obtained when 100% of ARs are verified as the most discriminant rules for classification are obtained (see figure 13.7). According to this, for the CTRL versus AD group, accuracy results versus the percentage of verified rules are shown in fig-



(a)



(b)

Figure 13.7: Accuracy results for the set CTRL versus MCI at 100% of *minconf* and *minsup* when FDG, PIB or the combination of both are used. a) CTRL-AD, b) MCI-NOR.AR-continuous-based method, biomarker database.

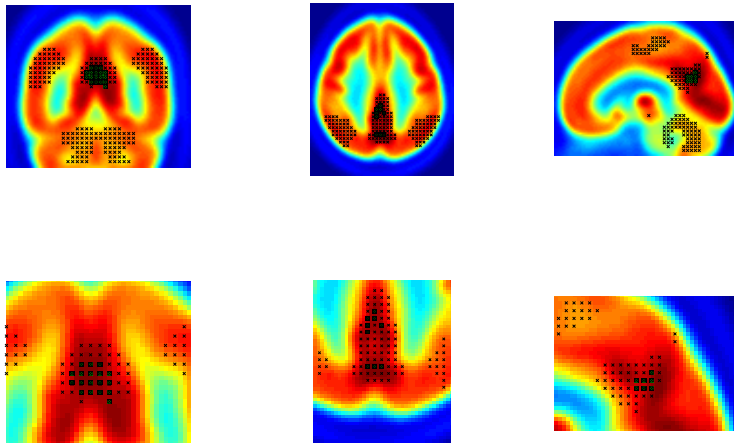


Figure 13.8: Bidimensional activated regions from coronal, axial and sagittal slices FDR-AR-based method. FDG-PET database.

ure 13.7(a) reaching maxima (Acc/Sen/Sp) rates of (92.11%/84.21%/100%) for FDG and (97.37%/94.74%/100%) for PIB biomarker. The combination without repetition of the ROIs obtained from both FDG and PIB biomarker databases for each subject, improves the FDG results as maxima rates of (94.74%/89.47%/100%) were obtained.

On the other hand, for the CTRL versus MCI group, accuracy results versus the percentage of verified rules are represented in figure 13.7(b). Maximum results (Acc/Sen/Sp) turned out to be better for PIB (90.48%/87.69%/100%) in comparison with FDG biomarker (77.38%/70.7%/100%) showing the same tendency than for the CTRL versus AD set. However, the combination of *FDG + PIB* biomarkers is very positive in the evaluation of this set as maxima results of (92.85%/90.77%/100%) are obtained improving the ones with solely one biomarker. In comparison with other reported methods, AR-based proposed technique outperforms PCA in combination with SVM [Illán et al., 2010], in which maximum Acc of 86.11% was reached.

	$h = 0.2$	$h = 0.4$	$h = 0.6$	$h = 0.8$
PET: Max number of mined rules	7482	5256	1332	210
PET: Acc (%)	91.33	90.67	88.67	84.00
SPECT: Max number of mined rules	67860	132	0	0
SPECT: Acc (%)	92.68	65.98	-	-

Table 13.2: Comparison of maximum Accuracy and number of AR mined obtained for SPECT and PET databases for different thresholds of the FDR voxel selection mask.

13.6.2 Experiments of AR over discriminant regions

Several experiments were conducted in order to test the reliability of the proposed FDR-AR-based classifier that consists in a FDR masking process (described in section 13.6.1) and an AR-based classification (explained in section 13.2.3). Two image modality databases (SPECT and PET) were used in the analysis in order to highlight its generalization ability. ROIs were selected as $5 \times 5 \times 5$ -voxel 3D blocks located inside the corresponding SPECT or PET FDR mask (explained in section 13.3.1 and depicted in figures 13.1.b and 13.1.c). An FDR threshold $h = 0.2$ was selected. The center coordinates of the 3D blocks are restricted to be in $2 \times 2 \times 2$ 3D grid. At the feature extraction stage, a 3D block is defined as activated if more than a fraction (50 voxels) of the total (125 voxels) are activated. This threshold represents a 40% of activated voxels and allows only brain volumes with a high activation to be considered as ROIs. For both databases, the activation threshold is reduced when the FDR mask is used since it locates itself the most discriminant voxels for the early detection of the AD.

Figure 13.8 shows coronal, transaxial and sagittal slices of a mean control PET image illustrating the voxels within the FDR mask (crosses) as well as the activated ROIs representing the ARs (squares), which are enlarged in the lower part of the figure 13.8 for clarity. Note that, these ROIs correspond to those regions commonly affected by AD, i.e., the posterior cingulate gyri and precuneus, as well as the temporo-parietal region [López et al., 2011]. Finally, the trade-off between computational cost and detection ability was studied in order to analyze the computational demands of the proposed method. Table 13.2 shows the computational cost of the proposed system in terms of the number of ARs mined for a given FDR threshold h . Note that, this threshold is used as a voxel selection criterion and enables reducing the dimension of the input variables in the AR mining process. The accuracy of the system improves by reducing the threshold h at the cost of increasing the number of

mined ARs and, subsequently, the computational complexity. The best trade-off between classification results and complexity is obtained for a threshold $h = 0.2$.

The accuracy, specificity and sensitivity values converge to a maximum of 92.78%, 100% and 87.5% for SPECT and 91.33%, 100% and 82.67% for PET, respectively. The classification accuracy improves when increasing the threshold of verified rules by each subject since a higher number of rules are required to be fulfilled to classify any subject under test as normal, thus more AD patients are correctly detected. In particular, 67860 rules represent the maximum number of verified rules for SPECT, while for PET, this number reaches 7482 rules (see table 13.2). As a conclusion, the size of the area under study derived from the FDR analysis in SPECT is wider than in PET where functional deficits are more localized in well known brain regions. Thus, the number of ROIs used as inputs for the AR mining process is higher for SPECT because of the size of the activation map mask. It is also observed that the specificity values for this CAD system reach 100%, that is, all controls are correctly classified when the *minsup* and *minconf* values in the AR mining process are equal to 100%. These results are in agreement with the expected behaviour of the system since normal subjects are assumed to have a common image pattern and to verify most of the mined ARs. Finally, it can be concluded that the proposed system outperforms recent developed CAD systems based on VAF and PCA [Illán et al., 2011b] that yield lower image classification accuracies of 83% or 86% for SPECT and 81.18% or 88.24% for PET, respectively. Figure 13.9 shows the ROC curves of the proposed FDR AR-based CAD system as well as baseline VAF and PCA methods for the SPECT and PET databases. The proposed method at 100% *minconf* and *minsup* outperforms several recently reported CAD systems, i.e. VAF or PCA [Illán et al., 2011b], since its ROC is shifted up to the left in the ROC space. In fact, the proposed method yields 87.5% and 82.67% sensitivity for SPECT and PET modalities, respectively, and 100% specificity. Figure 13.9 also shows the ROC curve of the proposed method for 97% *minconf* and *minsup*. In this case, the AR-based method reports lower accuracy rate than the previous model configuration, however this analysis provides useful information for parameter selection.

13.6.3 Experiments of discretization-AR-based system

The proposed system that consists of four stages: *i*) masking procedure, *ii*) discretization of mean intensity for feature selection, *iii*) AR mining, *iv*) loo cross-validation.

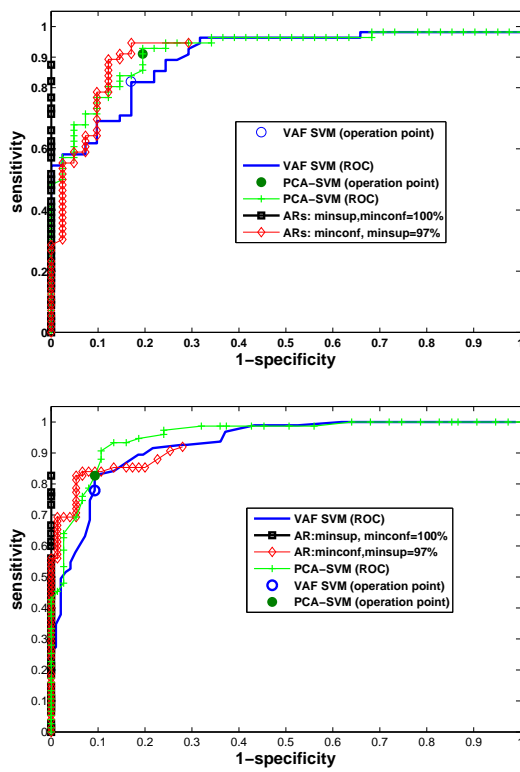


Figure 13.9: ROC Curve for FDR-AR-based method a) SPECT and b) PET databases. Comparison to other recently reported methods.

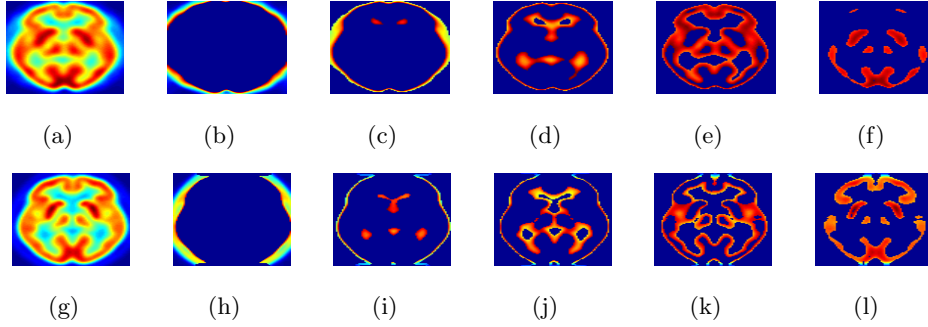


Figure 13.10: Transaxial slice of the image histogram segmentation (SPECT in the first row and PET in the second one) obtained from control average images (depicted in (a) and (g)) which are respectively divided in the following $K_1 = 5$ subimages corresponding to each bin. Discretization-AR-based method.

Masking for discretization procedure

First of all, the histogram is calculated over a mean control image to segment it into K_1 subimages that consist of the voxels whose intensities belong to each bin. This is also called image thresholding as it is based on the assumption that objects can be distinguished by their intensity level [Nakib et al., 2007]. The mask discards the voxels of the subimage with a value of zero (represented in blue in figure 13.10).

Figure 13.10 depicts (in the first row for SPECT and in the second one for PET) the transaxial slice masks obtained from the average of control images which are represented in figures 13.10.a and 13.10.g respectively for both databases when the $K_1 = 5$ subimages (from 1 to 5 in figures 13.10.b-f and 13.10.h-l) are segmented by means of an intensity histogram. It is observed that the selected masks (the 4th for SPECT depicted in figure 13.10.e and the 5th for PET in figure 13.10.l) include the most relevant voxels for the AD diagnosis since they are situated in the temporoparietal regions.

Voxels selected by discretization

The mask coordinates are the center of the 3D $v \times v \times v$ -sized cubic blocks. Each block is represented by a continuous feature that consists of the mean intensity (I_{mean}) of the voxels whose intensity belongs to the mask range $[I_{min}, I_{max}]$. The discretization consists of an equal-width-size histogram applied to the range $[I_{mean-min}, I_{mean-max}]$ dividing it into K_2 bins of equal width. Discretized bins are used independently or in a combined way to de-

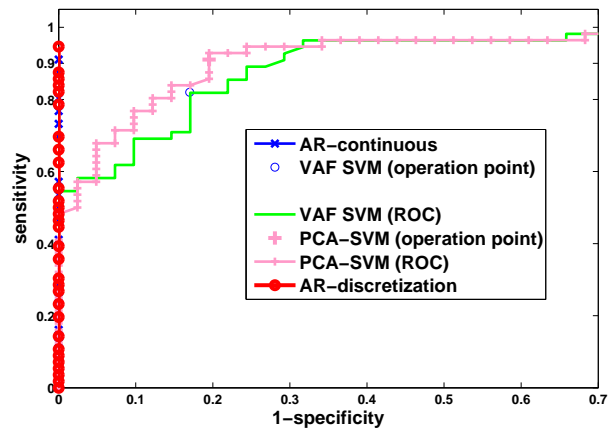
Bins of discretization	1	2	3	4	5
1	-/-	44.33/-	64.95/86.67	96.91/91.33	96.91/78
2	44.33/-	44.33/-	64.95/86.67	51.55/57.33	96.91 /
3	64.95/86.67	64.95/86.67	79.38/88	74.23/64	67.01/59.33
4	96.91/91.33	51.55/57.33	74.23/64	96.91/91.33	95.88/ 92
5	96.91/78	96.91/78	67.01/59.33	95.88/ 92	96.91/78

Table 13.3: Accuracy rate (%) taking 4th bin subimage ($K_1 = 5$) as mask and combining different bins of discretization stage ($K_2 = 5$) for SPECT/PET databases.

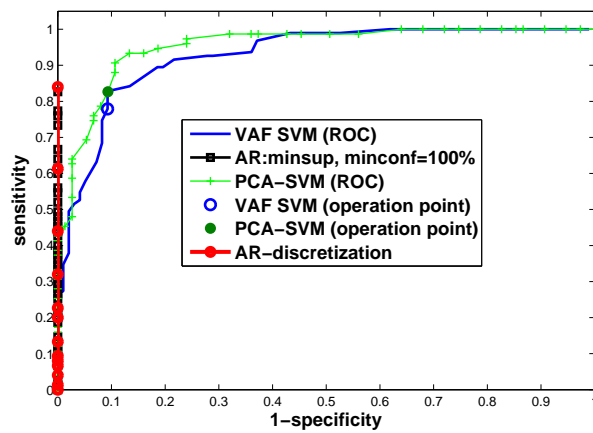
fine the intervals of selected ROI value for each subject. These ROIs act as inputs for the AR-mining (see section 13.2.1). It is demonstrated that while lower discretized bin combinations use to be noisy for classification, higher ones are likely to show relevant information for the AD early detection (see table 13.3).

Table 13.3 shows the discretized bin combinations when $K_2 = 5$ and it is demonstrated that when the higher bins are included in the combination, accuracy results are likely to improve. In fact, for SPECT when the 5th discretized bin is combined with the 1st, 2nd or 5th, 96.91% of accuracy is obtained (94.64% sensitivity and 100% specificity), outperforming the continuous AR-based technique [Chaves et al., 2011a] with results ($Acc = 94.87\%$, $Sen = 91.07\%$, $Spe = 100\%$). For PET in table 13.3, it is shown discretized bin combinations when the mask presented in figure 13.10.1 is applied, reaching rates of $Acc = 92\%$, $Sen = 84\%$ and $Spe = 100\%$ (when the 4th and the 5th bins are combined) outperforming the AR-based continuous case ($Acc = 91.33\%$, $Sen = 82.6\%$, $Spe = 100\%$) [Chaves et al., 2012c].

Figure 13.11 shows the ROC curves of the discretized-AR-based CAD system for SPECT (figure 13.11.a) and PET (figure 13.11.b). ROC curve for the proposed method is depicted at the best results of the discretized bin combinations (reported in table 13.3). As shifted up and to the left, this work outperforms other reported techniques such as PCA or VAF in combination with SVM or the AR-continuous based method [Chaves et al., 2011a] of section 13.6.1. Note that, as the $minconf$ and $minsup$ are selected at the maximum 100% (for continuous and discretized AR-based methods), consequently all controls are correctly classified ($Spe = 1$ in the ROC space) since the ARs obtained are the most discriminant in order to distinguish better normal from AD pattern.



(a)



(b)

Figure 13.11: ROC Curve a) SPECT and b) PET databases. Comparison with AR-baseline for continuous features and other reported methods: VAF-SVM and PCA-SVM.

	PCA: Acc/Sen/Spe (%)	PLS: Acc/Sen/Spe (%)	AUC: PCA/PLS
AR (90% AE, 100% minconf, minsup)	89.69/87.5/92.68	90.72/89.29/92.68	0.9549/0.9259
AR (90% AE, 95% minconf, minsup)	87.62/85.71/90.24	88.66/91.07/85.37	0.9017/0.9011
AR (90% AE, 90% minconf, minsup)	84.53/83.93/85.37	86.60/87.50/85.37	0.8915/0.8924
AR (92% AE, 100% minconf, minsup)	85.56/82.14/90.24	87.63/83.93/92.68	0.9277/0.9150
t-test	81.44/82.14/80.49	85.57/87.50/82.93	0.9317/0.9029
wilcoxon	79.38/80.36/78.05	89.69/91.07/87.80	0.9205/0.9230
energy	53.61/92.86/0	60.82/92.86/17.07	0/0

Table 13.4: Comparison of Acc/Sens/Spe rate and AUC for different FE techniques with final reduction of 6 features (SPECT database. FS-AR based method.)

13.6.4 Experiments of FS-AR-based system

The proposed system consists of four stages: *i*) preprocessing, *ii*) voxel selection, *iii*) feature extraction, and *iv*) SVM classification. The preprocessing stage performs spatial and intensity normalization of SPECT or PET data. As in the experiments of section 13.6.1, in the first step (preprocessing) and after functional images are normalized, a 3D mask is computed by averaging all control images from the database. An activation threshold intensity a_T is fixed by the 50% of the maximum intensity of this mean image, defining a 3D mask.

Voxel selection is enabled by AE and AR-mining techniques described in sections 13.2.2 and 13.2.1 respectively. Voxels are selected by AR combination stage, that is, the antecedents and consequents of previously obtained ARs are merged without repetition and the output of the AR-enabled voxel selection method is the activation level of these ROIs as explained in section 13.5. Then PCA or PLS are used to further reduce the dimension of the input feature vector to a kernel SVM.

Experiments on SPECT database

The center coordinates of the 3D blocks $9 \times 9 \times 9$ sized are restricted to be in $4 \times 4 \times 4$ 3D grid. At the activation estimation stage, a 3D block is defined as activated if more than a fraction (650 voxels) of the total (729 voxels) are activated. This threshold t represents a percentage above 90% of activated voxels, since it allows only brain volumes with a high activation to be considered as input for the AR-mining.

Figure 13.12 shows the fraction (%) of AR-based selected ROIs from the

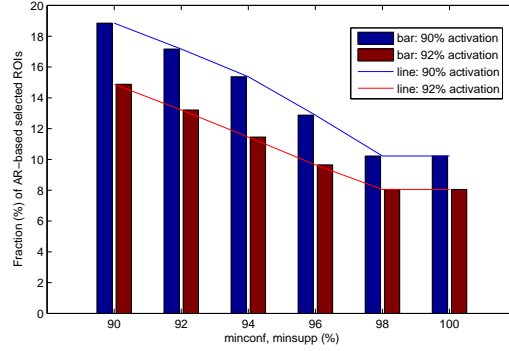
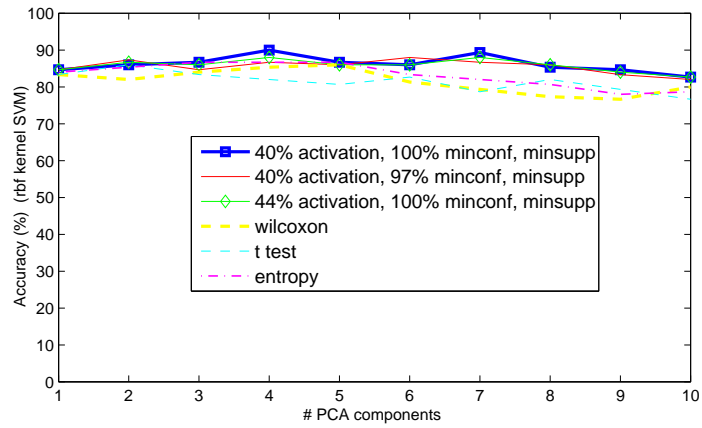


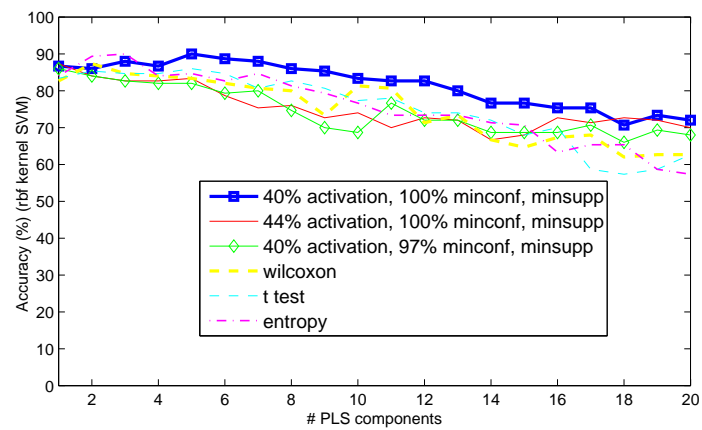
Figure 13.12: Fraction (%) of AR-based selected ROIs versus $minsup$ and $minconf$ (%) for different parameters: Activation threshold and $minconf/minsup$. SPECT database. FS-AR based method.

total of 2445 activated ROIs when percentages of 90% and 92% of activation are used. The number of rules is progressively reduced in a similar way to a negative exponential distribution from 90% to 100% of $minsup$ and $minconf$. While at 90% of $minsup$ and $minconf$, the fraction (%) of AR combination selected from the total of ROIs reaches 18.85% and 14.88% (for activation of 90% and 92%), at 100% this number appears reduced to 10.22% or 8.05% respectively. Despite decreased the number of merged ARs as the activation level increases above 90%, this provokes the losing of some discriminant rules for the AD early diagnosis in the light of the results (see table 13.4 or figure 13.14 for further details) when 92% of activation is used. Although this reduction is considerable, a subsequent PCA or PLS feature extraction is necessary to improve the computational time and accuracy rate of the system.

Table 13.4 shows the Acc, Sen and Spe of the proposed AR-based FS method and comparison with other FS baseline techniques at 6 features (Variance explained accounts for the variation of a feature subset when PCA or PLS strategies are applied and in this case it does not change significantly above 6 components). Acc, Spe, Sen values converge to 89.69%, 87.5% and 92.68% for PCA and 90.7%, 92.68% and 89.29% for PLS when a $minsup$ and $minconf$ of 100% and an AE threshold of 90% are used, which outperform the other baseline FS-based methods.



(a)



(b)

Figure 13.13: Accuracy, sensitivity and specificity values as a function of the number of components for PET database: a) PCA, and b) PLS. FS-AR based method.

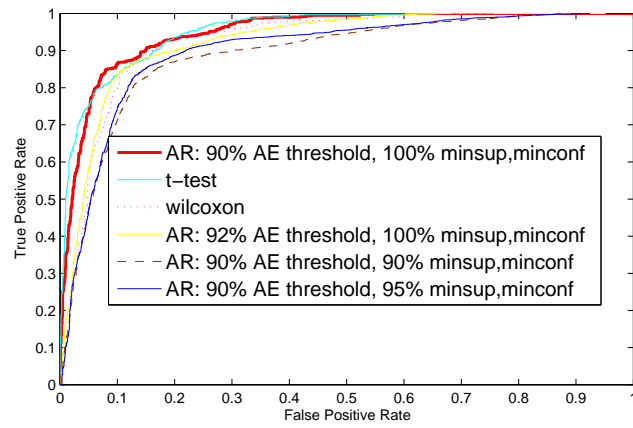
Experiments on PET database

While voxel intensity is related to blood flow in SPECT imaging, for PET image modality the activation level is related to glucose metabolism. Generally speaking, PET images are experimentally found to be less activated than SPECT images in each ROI. For this reason 3D-Blocks are selected in these experiments with a smaller size, i.e. $5 \times 5 \times 5$ -voxels per block, with center coordinates restricted to be in a $2 \times 2 \times 2$ grid. A voxel activation above 40% allows including relevant ROIs in the AR-based feature selection.

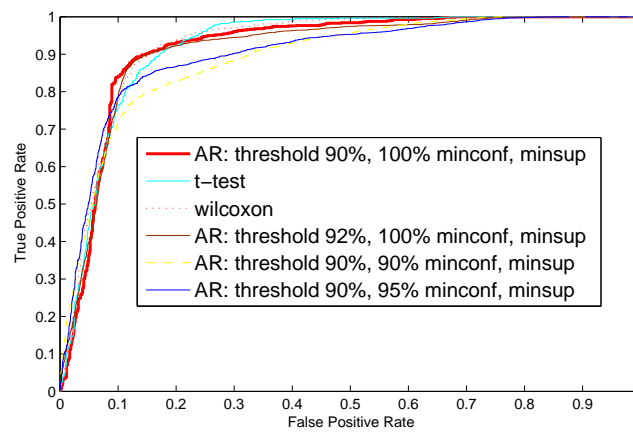
In a similar way to SPECT modality, the number of AR-based selected ROIs decreases more at 100% of *minconf* and *minsup* from the total of 3776. In particular, the fraction of AR-based selected ROIs is 4.87% or 6.09% when activation is of 44% or 40% respectively. Although at 44% of activation the reduction is higher, the system discards some relevant ROIs as shown in accuracy results of figure 13.13. The Acc, Sen and Spe are estimated by means of k -fold cross-validation with a rbf SVM classifier as it performs better for PET modality. Figures 13.13.a and 13.13.b considered an increasing number of PCA or PLS features respectively yielding the maximum rates Acc, Sen and Spe of 90%, 85.33%, 94.67% at 4 PCA features or 90%, 89.33%, 90.67% at 10 PLS features respectively that outperform other reported techniques as t-student, Mann-Whitney-Wilcoxon test and Relative Entropy-based.

ROC analysis

Figure 13.14 and 13.15 show the ROC curves of the proposed AR-FS-based CAD system for respectively the SPECT and PET modalities considering the PCA (figures 13.14.a and 13.15.a) and PLS (figures 13.14.b and 13.15.b) FE techniques. The proposed AR FS-based method outperforms the baseline techniques as well since its ROC is shifted up to the left in the ROC space. This improvement of the proposed method is also demonstrated by the AUC, which is detailed in table 13.4 for SPECT yielding values of 0.9549 and 0.9259 for PCA and PLS modalities, respectively at 100% *minconf* and *minsup* and 90% AE threshold. As PET database is concerned, the AUC shows a trend (detailed in caption of figure 13.15) similar to SPECT, that is, it yields the highest results of 0.9344 and 0.9391 for PCA and PLS modalities at 100% *minconf* and *minsup* and 40% AE threshold.

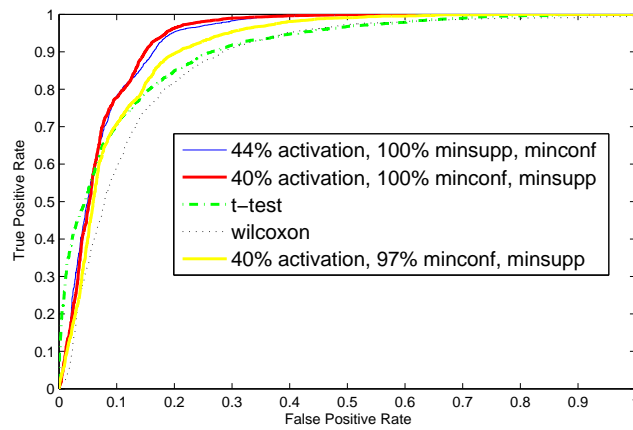


(a)

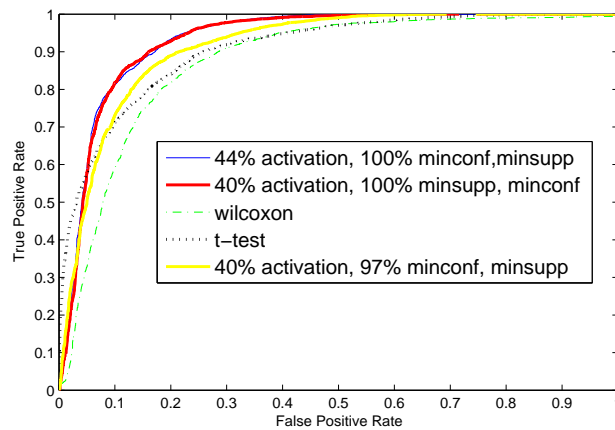


(b)

Figure 13.14: ROC Curve for SPECT database a) PCA and b) PLS. Comparison to other recently reported methods. FS-AR based method.



(a)



(b)

Figure 13.15: ROC Curve for PET database a) PCA and b) PLS. Comparison to other recently reported methods. The AUC obtained for each ARFS-based ROC (PCA/PLS) is: 100% minconf and minsupp with 40% activation (0.93441/0.9391), 44% activation (0.9107/0.9181) and 97% minsupp and minconf with 40% activation (0.9107/0.9181). AUC (PCA/PLS) for other reported methods is: wilcoxon (0.8662/0.8662) or t test (0.8968/0.9027).

Chapter 14

Large margin nearest neighbors

In this chapter, after an introduction to the large margin principle and its loss function, LMNN is described from different points of view: i) as a transformation of the feature space obtained by means of PLS or PCA in order to better separate the control subject and AD patient classes, ii) as feature reduction technique by a rectangular matrix (LMNN-RECT), and iii) as a classifier comparing Euclidean, Mahalanobis and energy-based methods. Both i) and ii) make use of a kernel-SVM and the Kernel LMNN methods. Stability and robustness of this LMNN-based method is shown by means of experiments made in SPECT and FDG-PET (ADNI) databases.

14.1 Introduction

In the last few years, the large margin principle has been used to design classification algorithms [Hu et al., 2011]. For example, the nearest neighbor classification is a simple and yet effective technique for pattern recognition. Performance of this technique depends significantly on the distance function (typically the Euclidean one) used to compute similarity between examples. Every unseen point x is compared through a distance function $dist(x, x_i)$ to all points x_i ($i=1,2,\dots,n$) of the training set. The k minimal distances are computed and the majority over the corresponding labels y_i is taken as the resulting label for x [S. Lemm, 2011].

In recent years, there has been growing interest in finding variants of KNN [Goldberger et al., 2005]. In [Weinberger and Saul, 2008], a method for learning a Mahalanobis distance metric from training samples by semidefinite programming was introduced. In addition, margin based loss functions, instead of margin itself, are used in characterizing the complexity of classification. Classification loss is a function of margin, such as hinge loss used in SVM. Breiman showed that AdaBoost minimizes the exponential loss function of margin [Breiman, 1999]. Weinberger et al. proposed LMNN for k nearest neighbor classification from labeled examples using hingeloss.

LMNN is used in this work in different ways: *i*) as a transformation of the feature space obtained by means of PLS or PCA in order to better separate the control subject and AD patient classes, *ii*) as feature reduction technique by performing the transformation as a rectangular matrix (LMNN-RECT), and *iii*) as a classifier as reported in section 14.1.4. The objective of LMNN is to obtain a family of metrics over the feature space. Let $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ denote a training set of n labeled examples with inputs $\mathbf{x}_i \in \mathbb{R}^d$ and associated class labels \mathbf{y}_i . Our goal is to learn a linear transformation $\mathbf{L}: \mathbb{R}^d \rightarrow \mathbb{R}^d$. These metrics compute squared distances as:

$$\mathbf{D}_{\mathbf{L}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2^2 \quad (14.1)$$

Equation 14.1 is commonly used to express squared distances in terms of the squared matrix:

$$\mathbf{M} = \mathbf{L}^T \cdot \mathbf{L} \quad (14.2)$$

On the other hand, the squared distances are denoted as Mahalanobis metrics in terms of \mathbf{M} :

$$\mathbf{D}_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \cdot \mathbf{M} \cdot (\mathbf{x}_i - \mathbf{x}_j) \quad (14.3)$$

A Mahalanobis distance can be parameterized in terms of the matrix \mathbf{L} or the matrix \mathbf{M} [Weinberger et al., 2009]. The first is unconstrained, whereas

the second must be positive semidefinite. The main idea of LMNN consists of minimizing the loss function (see the following section 14.1.1) that is able to learn a distance metric under which inputs and their target neighbours are closer together.

14.1.1 Loss function

In LMNN, target neighbours are defined as input patterns of the same class that are wanted to be closer. The loss function to be minimized consists of two terms. One acts to *pull* target neighbours closer together penalizing large distances between each input and its target neighbours. The other term acts to *push* differently labeled examples further apart. It penalizes small distances between differently labeled examples. The *pull* term is represented by the following equation:

$$\varepsilon_{pull}(\mathbf{L}) = \sum_{j \rightarrow i} \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 \quad (14.4)$$

where $j \rightarrow i$ means that input \mathbf{x}_j is a target neighbour of input \mathbf{x}_i . A new indicator variable is introduced to define the *push* term of the loss function:

$$\mathbf{y}_{il} = \begin{cases} 1 & \text{if } \mathbf{y}_i = \mathbf{y}_l \\ 0 & \text{otherwise} \end{cases} \quad (14.5)$$

so that:

$$\varepsilon_{push}(\mathbf{L}) = \sum_{i,j \rightarrow i} \sum_l (1 - \mathbf{y}_{il}) \cdot [1 + \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 - \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|^2]_+ \quad (14.6)$$

where $[\]_+ = \max(z, 0)$ denotes the standard hinge loss [Weinberger et al., 2009]. Finally, we combine the two terms $\varepsilon_{pull}(\mathbf{L})$ and $\varepsilon_{push}(\mathbf{L})$ into a single loss function for distance metric learning. The two terms can have competing effects, to attract target neighbours and to repel impostors. Impostors are defined as the inputs with different labels. A weighting parameter $\mu \in [0, 1]$ balances these goals.

$$\varepsilon(\mathbf{L}) = (1 - \mu) \cdot \varepsilon_{pull}(\mathbf{L}) + \mu \cdot \varepsilon_{push}(\mathbf{L}) \quad (14.7)$$

14.1.2 LMNN-RECT as feature reduction technique

The loss function needs to be optimized in order to obtain the distance metric transforms in terms of the explicitly low-rank linear and rectangular matrix transformation \mathbf{L} . The optimization over \mathbf{L} is not convex unlike the original optimization over \mathbf{M} , but a (possibly local) minimum can be computed

by standard gradient-based methods. We call this approach LMNN-RECT [Weinberger and Saul, 2008], in which \mathbf{L} is a matrix with a size equal to the number of features selected by the t -test. In particular, in this work the matrix \mathbf{L} is multiplied by the matrix consisting of the NMSE features selected by the t -test and defined above in order to obtain a new space of features that better separates control subjects from AD patients. This fact is experimentally demonstrated in the Experiments section 15.1.

14.1.3 Kernel LMNN

It is interesting to consider the case where \mathbf{x}_i are mapped into a high dimensional feature space $\phi(\mathbf{x}_i)$ and a Mahalanobis distance is sought in this space. We focus on the case where dot products in the feature space may be expressed via a kernel function, such that

$$\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j) \quad (14.8)$$

for some kernel k [Globerson and Roweis, 2005]. When we use the Kernel PCA trick framework, the original LMNN can be immediately used as Kernel LMNN (KLMNN) as it is explained in [Chatpatanasiri et al., 2010]. The new KPCA trick framework offers several practical advantages over the classical kernel trick framework, e.g. no mathematical formulas and no reprogramming are required for a kernel implementation, a way to speed up an algorithm is provided with no extra work, the framework avoids troublesome problems such as singularity.

14.1.4 Classification:

LMNN and SVM classifiers were used in this work to build the AD CAD system. They present many similarities, for example its potential to work in nonlinear feature spaces by using the kernel trick. On the other hand, features can be extracted by means of the kernel trick and PCA (kernel PCA, KPCA) or LMNN (kernel LMNN, KLMNN) [Chatpatanasiri et al., 2010]. LMNN can be viewed as the logical counterpart to SVMs in which kNN classification replaces linear classification. However, LMNN contrasts with classification by SVMs, in that it requires no modification for multiclass problems that involve combining the results of many binary classifiers, that is there is no explicit dependence in the number of classes.

LMNN classifier

Some techniques were developed to learn feature weights to manage the change of distance structure of samples in nearest neighbour classification. Euclidean distance, the most commonly used, assumes that each feature is equally important and independent from others. By contrast, a distance metric with good quality such as Mahalanobis, should identify relevant features assigning different weights or importance factors to the extracted ROIs [Xiang et al., 2008]. Only when the features are uncorrelated, the distance under a Mahalanobis distance metric is identical to that under the Euclidean distance metric. On the other hand, our work has been inspired by energy-based metric (EBC) learning, obtaining with it the best results in terms of accuracy, specificity and sensitivity [Chopra et al., 2005; Pérez et al., 2000]. EBC consists of computing the loss function for every possible label \mathbf{y}_i . We compute the minimization of three terms. The first one term is defined to be the squared distances to the k target neighbours of \mathbf{x}_i . The second term accumulates the hinge loss over all impostors (that is differently labeled) which invade the perimeter around \mathbf{x}_i determined by its target neighbours. The third term is the accumulation of the hinge loss for differently labelled examples whose perimeters are invaded by \mathbf{x}_i .

SVM classifier

SVMs [Vapnik, 1998] let to build reliable classifiers in very small sample size problems and even may find nonlinear decision boundaries for small training sets. SVM [Burges, 1998] separates a set of binary-labeled training data by means of a maximal margin hyperplane, building a decision function $\mathbb{R}^N \rightarrow \{\pm 1\}$. The objective is to build a decision function $\mathbf{f}: \mathbb{R}^N \rightarrow \{\pm 1\}$ using training data that is, l N -dimensional patterns \mathbf{x}_i and class labels y_i : (\mathbf{x}_1, y_1) , (\mathbf{x}_2, y_2) , . . . , (\mathbf{x}_l, y_l) , so that \mathbf{f} will correctly classify new unseen examples (\mathbf{x}, y) .

Linear discriminant functions define decision hyperplanes in a multidimensional feature space: $\mathbf{g}(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + w_0$ where \mathbf{w} is the weight vector to be optimized that is orthogonal to the decision hyperplane and w_0 is the threshold. The optimization task consists of finding the unknown parameters w_i , $i = 1, \dots, N$ and w_0 that define the decision hyperplane. When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques such as quadratic, polynomial or radial basis function (RBF), so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space [Bishop, 2006].

***k*-fold cross-validation evaluation**

Although there are studies that consider k independent training and test splits (for instance in [Nadeau and Bengio, 2003] and [Bengio and Grandvalet, 2004]), we focus on the standard k -fold cross-validation that is widely used ([Westman et al., 2011], [Illán et al., 2011a], [Bengio and Grandvalet, 2004]). In k -fold procedure, there is no overlap between test sets: each example of the original data set is used once and only once as a test example. In k -fold cross-validation, sometimes called rotation estimation, the dataset D is randomly split into k mutually exclusive subsets (the folds) D_1, D_2, \dots, D_k of approximately equal size. The inducer is trained and tested k times; each time $t \in \{t_1, t_2, \dots, t_N\}$, it is trained on $D - \{D_t\}$ and tested on D_t [Kohavi, 1995].

10 folds were used in each experiment which yielded accurate estimates of the error rates. For each iteration ($t = 1, \dots, 10$), the algorithm returns randomly generated indices for a k -fold cross-validation of D observations. Testing rate is mostly equal to the integer of the fraction $100/\text{number of folds}$, that is 10% in our experiments, but it can vary randomly one or two samples in each iteration if the number of observations is a prime number. These indices are used for testing and the rest (approximately 90%) for training. Statistical results obtained in each iteration are averaged. Thus, by using cross-validation, several feature extraction and classification methods were objectively compared in terms of their respective fractions of misclassified samples. In this way, the classifier was evaluated in depth as a tool for the early detection of AD in terms of the Acc, Sen and Spe.

14.2 Experiments

Several experiments were conducted in order to evaluate the proposed LMNN-based feature extraction algorithms and its benefits as: *i*) linear transformation of the PLS or PCA reduced data, *ii*) feature reduction technique, and *iii*) classifier (with Euclidean, Mahalanobis or Energy-based methodology). SVM classification including transformation of the input space by means of linear, polynomial, quadratic or rbf kernels, which define non-linear decision surfaces, was adopted for the first two approaches. The classification performance of our approach was tested by means of k -fold cross validation (instead of loo), which is widely used to compare the performances of different predictive modeling procedures as in [Wiens et al., 2008].

Firstly, control subjects are averaged in a tridimensional image $sm(x, y, z)$. In functional imaging, each voxel carries a grey intensity level $I(x_j)$, which is related to the regional cerebral blood flow, glucose metabolism, etc. in the

brain of a patient, depending on the image acquisition modality. Secondly, it is obtained a 3D $\mathbf{mask}(x, y, z)$ that consists of all the voxels with $\mathbf{sm}(x, y, z) > a_T$. The threshold a_T is equivalent to the 50% of the maximum intensity in $\mathbf{sm}(x, y, z)$.

Baseline VAF is a way of including in $\mathbf{vaf}(x, y, z)$ all the voxels inside the obtained $\mathbf{mask}(x, y, z)$ and considering them as features. Therefore, voxels outside the brain and poorly activated regions are excluded from this analysis. The main problem to be faced up by these techniques is the well-known small sample size problem, that is, the number of available samples is much lower than the number of features used in the training step. However in this work, the combination of feature reduction techniques does not only solve this problem, but also helps to reach better results of classification.

Instead of using directly all the voxels, the regions are considered in 3D because not all the brain regions provide the same discriminant value for detecting the early AD. In fact, the posterior cingulate gyri and precunei, as well as the temporo-parietal region are typically affected by hypo-perfusion in the AD [Chaves et al., 2011a]. That is the reason why, each functional image is processed by means of 3D $v \times v \times v$ cubic voxels defining ROIs, or $\mathbf{block}(x, y, z)$ centered in (x, y, z) coordinates which belong to $\mathbf{vaf}(x, y, z)$. Then, it is calculated the Normalized Minimum Squared Error or $NMSE_p(x, y, z)$ defined as:

$$NMSE_p(x, y, z) = \frac{\sum_{l,m,n=-v}^v [f(x-l, y-m, z-n) - g_p(x-l, y-m, z-n)]^2}{\sum_{l,m,n=-v}^v [f(x-l, y-m, z-n)]^2} \quad (14.9)$$

It is obtained for each subject and block where $f(x, y, z)$ is the mean voxel intensity of all the control subjects and $g_p(x, y, z)$ is the voxel intensity of the p -th subject at (x, y, z) coordinates. The most discriminant ROIs are obtained by means of an absolute value two-sample t -test with pooled covariance estimate on NMSE features as in [Chaves et al., 2009b].

Widely used methods for the analysis of data sets are PCA [Jackson, 1991], [Wold et al., 1987] and projections to latent structures (PLS) [Wold, 1982], [Tenenhaus, 1998], that work computationally well for many variables and observations. By contrast, LMNN algorithm is aimed at the organization of the k -nearest neighbors to the same class, while examples from different classes are separated by a large margin [Weinberger et al., 2009], [Goldberger et al., 2005], [Chopra et al., 2005], [Tsochantaridis et al., 2005]. In this work we propose and compare several feature extraction methods that includes on the one hand the combination of NMSE feature extraction with PCA or PLS plus the LMNN transformation. On the other hand, NMSE is directly combined

with a LMNN-RECT reduction.

14.2.1 Experiments with SPECT

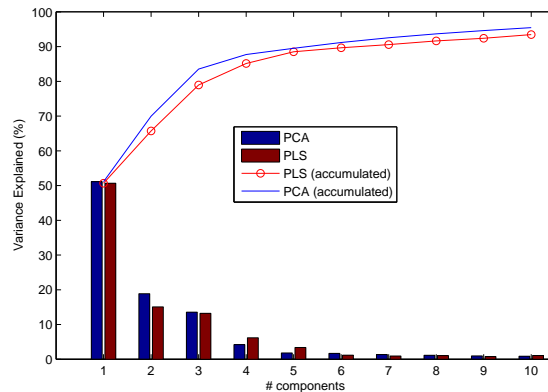


Figure 14.1: Variance Explained (%) versus PCA and PLS Components in bar diagram. Lines represent the accumulated Variance Explained (%) versus Principal Components and PLS Components.

The number of features used is a trade-off between ROIs that are really important and that do not worsen the computational time of the CAD. We demonstrated experimentally that 200 NMSE features (ROIs) is a number high enough to guarantee the quality of the image in posterior classification. However, this number must be reduced in order to improve the computational time of the system with strategies such as PCA, PLS or LMNN-RECT. The final number of features used has been experimentally tuned by the observation of figure 14.1 in which the percentage of variance explained for features (PCA or PLS) chosen are drawn as bars and a line represents the cumulative Variance Explained. In the case, Variance Explained accounts for the variation of a feature subset when PCA or PLS strategies are applied. In this graphic, we can observe that up to six components, the variance explained for PCA and PLS does not change significantly.

Finally, table 14.1 shows the accuracy, sensitivity and specificity rates of the proposed methods and compares them with other recently reported techniques including VAF, GMM and PCA. PCA and PLS feature extraction considered up to six features as they retain most of the variance of the data. It can be concluded from the table 14.1 that the LMNN transformation when

SVM-linear classifier	Accuracy (%)	Sensitivity (%)	Specificity(%)
VAF	83.51	83.93	82.93
PCA	86.56	91.07	80.49
GMM	89.69	90.24	89.29
Gaussian kernel PCA+LMNN Transformation	91.75	91.07	92.68
Gaussian kernel PLS+LMNN Transformation	90.72	91.07	90.24
PLS+LMNN Transformation	92.78	91.07	95.12
LMNN-RECT	80.28	70	87.80
LMNN-Classifier Accuracy (%)	Euclidean	Mahalanobis	Energy
PCA	80.54	81.63	87.65
PLS	84.33	89.56	88.67

Table 14.1: Statistical measures of performance of LMNN-based techniques in comparison with other reported methods for SPECT database. SVM classifier: Comparison of the methods reported in this work with VAF, GMM and PCA operation points. LMNN-based techniques parameters: linear SVM classifier with 6 components. VAF parameters: linear SVM classifier, GMM parameters: $\sigma=6$ RBF-SVM classifier with 8 components and PCA parameters: $\sigma=6$ RBF-SVM classifier with 16 components. LMNN Classifier with 6 components: Euclidean, Mahalanobis and Energy.

combined with PCA or PLS yields the best results and reports benefits when compared to other reference methods.

To sum up, LMNN was presented as a valid solution to make broader the margin between the classes. It was developed an effective CAD system in which it is not necessary to incorporate an *a priori* knowledge about the pathology, since up to its first feature extraction step, all the voxels with a considerable activation (that is, those voxels that are located inside the calculated mask) are considered. The analysis shown in this papers reports clear advantages in the following ROI-selection steps as well, because they were computed in an automatic way for the early diagnosis of Alzheimer's disease. The best combination of feature reduction techniques yielded an accuracy value of 92.72%, thus outperforming other recently and consolidated reported methods such as VAF, PCA and GMM (see table 14.1). Finally, in order to study in depth the AD classification with LMNN-based techniques, we have also included additional information about the classification of AD1 subjects versus CTRL. This set up is more difficult to be classified since AD1 pattern is still a challenge to be diagnosed. If we only consider the case CTRL versus AD1 the precision rates of the method are for PCA plus LMNN: Acc=84.51%, Sen=73.33%, Spe=92.68%, for PLS plus LMNN transformation: Acc=83.10%, Sen=70%, Spe=92.68% and for LMNN-RECT: Acc= 84.51%, Sen= 76.67%, Spe= 90.24%. These results still represent a great advance in the field in comparison with the baseline VAF: Acc= 77.46%, Sen= 70%, Spe= 82.93%.

Classification using LMNN using Euclidean, Mahalanobis and Energy-based models is analysed when NMSE-PCA or NMSE-PLS features are considered. Figures 14.2(a) and 14.2(b) show the accuracy obtained by LMNN classification as a function of the number of PCA coefficients and PLS coefficients, respectively. The results show that LMNN classification using energy-based models and Mahalanobis distances performs better than when the Euclidean distance is considered, which suffers a decrease in the accuracy as the number of features increases. LMNN classification using energy-based models and Mahalanobis distances were found to be very robust against the selection of the dimension of the feature vector yielding peak values of the accuracy of 91.76% and 90.87%, respectively, when NMSE-PCA features are used. If PLS technique is used instead of PCA, the accuracy results improved yielding accuracy values of 91.78% and 89.78% for energy-based models and Mahalanobis distance, respectively. In all these cases, energy-based models outperformed the others.

14.2.2 Experiments with PET

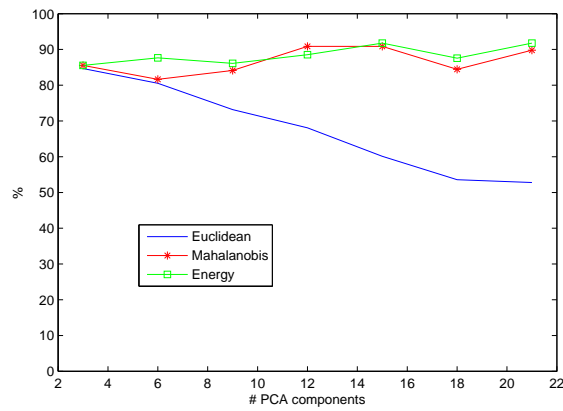
Additionally, several experiments were performed on a PET database in order to highlight the generalization ability of the proposed method. The same parameters such as voxel size or number of NMSE Blocks than for SPECT data were used. Figure 14.3(a) shows the different feature extraction techniques of this work, that is, PCA or PLS plus LMNN transformation and LMNN-RECT in comparison with PCA or VAF baseline when a linear SVM classification is performed. In the light of the graphic, this manuscript technique reaches a maximum accuracy rate of 90.67% (88% sensitivity and 93.33% specificity) for both PCA and PLS plus LMNN transformation and when used LMNN-RECT, accuracy 87.33% (82.67% sensitivity and 92% specificity), thus outperforming the PCA (Acc=85.33%) or baseline VAF (Acc=81.18%) techniques. In this work the advantage of the combination of the methods proposed, both for SPECT (see [Chaves et al., 2012d] for further details) and PET is that they maintain stable around the 90%. We can conclude that the fact of obtaining the ROIs by using the combination of NMSE Blocks with LMNN algorithm favors the stability in all the range of reduced features, thus promoting the robustness of the algorithm. Figure 14.3(b) shows LMNN classification using energy-based models, Mahalanobis and Euclidean distances for PCA and PLS features. Maxima accuracy rates were obtained for Energy-based classifier (90.11% for PCA and 89.99% for PLS).

14.2.3 ROC analysis

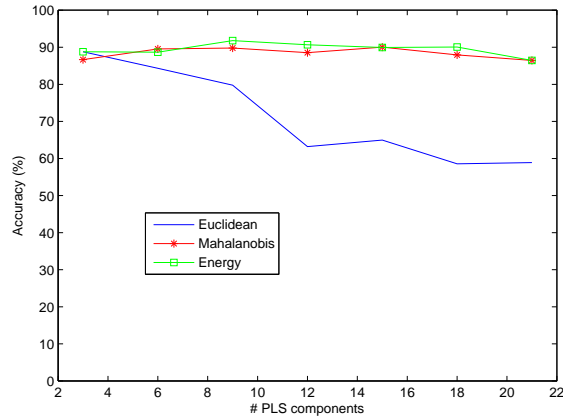
Figures 14.4(a) and 14.4(b) show the ROC curves of the proposed systems and other methods that were considered as a reference for SPECT and PET databases respectively. Several experiments were carried out on the different image modalities (SPECT and PET) in order to highlight the generalization ability of the proposed method. The analysis shows that the presented CAD system based on LMNN algorithm and SVM yields the best trade-off between sensitivity and specificity by shifting the operating point up and to the left in the ROC space [Metz, 1978] in comparison with other reported methods such as VAF SVM, PCA SVM and GMM SVM.

As shown in figure ROC analysis, PLS plus LMNN transformation provides an operation point located in the upper left than other approaches on the ROC space. In addition, the improvement of the proposed LMNN-based technique is also supported by the AUC analysis for SPECT/PET databases respectively: PLS plus LMNN transformation: 0.9424/0.9437, PCA plus LMNN transformation 0.9411/0.9505, LMNN-RECT: 0.9076/0.9325 that outperform

the AUC of other reported methods such as VAF SVM: 0.8993/0.8500 and PCA SVM: 0.9177/0.9006.

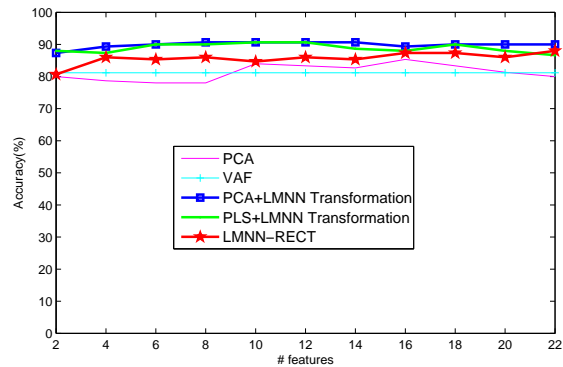


(a) PCA

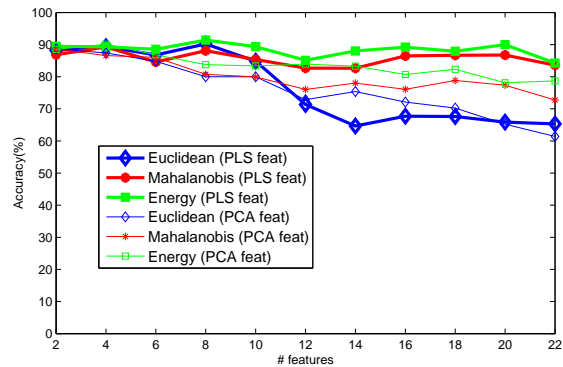


(b) PLS

Figure 14.2: LMNN classification (Euclidean, Mahalanobis and Energy-based models) for SPECT Feature reduction techniques: a) PCA, b) PLS

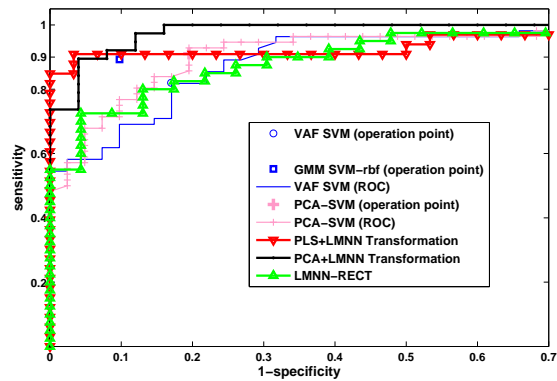


(a) SVM classifier

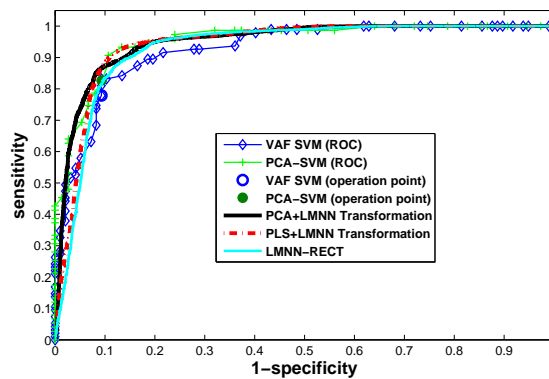


(b) LMNN-classifier

Figure 14.3: a) SVM classification: Accuracy, Specificity and Sensitivity (%) versus number of reduced features for PET database. Feature reduction techniques: PCA plus LMNN Transformation, PLS plus LMNN Transformation, LMNN-RECT, PCA and VAF b) LMNN classification (Euclidean, Mahalanobis and Energy-based models) for PCA and PLS features



(a) SPECT



(b) PET

Figure 14.4: ROC Analysis. LMNN-based methods SVM classified (PCA+LMNN Transformation, PLS+LMNN Transformation, LMNN-RECT). Comparison to other recently reported methods represented by their operation points. a) SPECT database. The AUC obtained for each ROC is: PCA+LMNN (0.9411), PLS+LMNN (0.9424), LMNN-RECT (0.9076), VAF SVM (0.8993) and PCA SVM (0.9177). b) PET database. The AUC obtained for each ROC is: PCA+LMNN (0.9505), PLS+LMNN (0.9437), LMNN-RECT (0.9325), VAF SVM (0.8500) and PCA SVM (0.9006).

Chapter 15

Conclusions and future work

This chapter shows briefly the conclusions, discussion and the main advantages and disadvantages of each proposed CAD system. Secondly, some future research lines for the continuation of this work are proposed.

15.1 Conclusions

CAD systems reported in this chapter can be divided in AR-based and LMNN-based and they all achieve a satisfactory classification performance over SPECT and PET images for neurological disease detection, in particular for the early detection of AD.

15.1.1 Conclusions of ARs

As the former is concerned, up to now AD has only been studied from a static point of view - usually by just looking for hypoperfusion patterns. By contrast, the innovation of this work consists in studying relationships among brain areas with ARs. In the light of the results this has turned out to be very efficient in a more dynamic way. Moreover, this CAD system was defined as an offline method, since the rules were obtained by means of the Apriori algorithm from a set of control subjects and every test subject was evaluated by means of the same set of rules.

ARs among continuous attributes

AR baseline classification depends on the verified (by each subject) number of rules, which were mined from controls. It was shown that Accuracy, Specificity and Sensitivity increase up to a maximum value as the threshold of verified ARs increases. However, this effect only happens when ARs are mined as being the most discriminative, with *minsup* and *minconf* values of 100%. The proposed method yields up to 94.87% of classification Accuracy (Sensitivity=91.07%, Specificity=100%) for SPECT data outperforming recent developed methods for the early Alzheimer's disease diagnosis. In any other case Accuracy and Specificity decrease at a determined number of rules, even when the most suitable rule for classification is chosen.

For the analysis of PET Cartuja database, ROIs with the same size (v, g) and with the same activation (t) than in SPECT have been used. As *minsup* and *minconf* are of 100%, the maximum rates (Acc=93.33%, Spe=100%, Sen=90.48%) are reached at 100% of verified rules, that is, 82082.

In addition, in this work, an AR-based classification method is applied over a PIB and [^{18}F]-FDG PET databases. A comparison is conducted in the sets AD versus CTRL and CTRL versus MCI. PiB is distinguished from FDG in providing differential information for the early diagnosis of AD, reaching maxima accuracies (Acc = 97.37% for CTRL versus AD and Acc = 90.48% for the CTRL versus MCI group). In the light of the results, it is concluded

that the combination of FDG and PIB biomarkers shows positive results for both groups: while in the first set FDG results are outperformed reaching $Acc = 94.74\%$, in the second set both FDG and PIB rates are improved reaching maxima of $Acc=92.86\%$.

ARs over discriminant regions

On the other hand, an FDR mask not only enabled selecting the most discriminant regions for further analysis with ARs and improving the classification results, but also reduced the computational cost (in comparison with the continuous case). Several experiments were conducted in order to test the reliability of the proposed AR-based classifier. Two medical image modality databases (SPECT and FDG-PET) were used in the analysis in order to highlight the generalization ability of the CAD system. The proposed method outperformed recently reported methods (i.e. VAF or PCA) in terms of accuracy, sensitivity and specificity and converged to a optimum operating point when the number of verified ARs was sufficiently high (including all the rules associated with reported brain regions that are relevant in early AD detection). In this way, it yielded an accuracy, specificity and sensitivity up to 92.78%, 100% and 87.5%, respectively, for SPECT, and 91.33%, 100% and 82.67%, for PET, when 100% minimum support and confidence were selected.

Discretization AR-based system

Integration of discretization for feature selection and AR for classification is proposed. In this way, we avoid the use of threshold for selecting activated ROIs, making the CAD working more automated. Firstly, an image histogram segmentation is applied over the mean control image in order to obtain the best mask, corresponding to higher bins (in particular the 4th for SPECT and the 5th for PET, from a total of K_1 subimages). The selection of $K_1 = 5$ for the mask obtaining implies a generalization in both databases and the assumption that the segmentation-based histogram is not misleading (which frequently occurs when too few or too many ranges are used). Secondly, an equal-width sized discretization is performed by the combination of different bins being the most activated (or higher) ones (from the total of K_2) usually the most suitable in order to choose the relevant 3D ROIs used as input in the AR mining. $K_2 = 5$ is selected as a trade-off between computational cost and accuracy results. The proposed method reaches $Acc = 96.91\%$ (for SPECT) and 92% (for PET) outperforming the AR-continuous baseline method described in [Chaves et al., 2011a] and other reported methods such as VAF-SVM or PCA-

SVM. We have shown classification results when used relevant discretized bins for AR-mining, with *minsup* and *minconf* of 100%. In that case, at the 100% of verified rules the maxima classification results are obtained and all controls are correctly classified, that is $Spe = 100\%$. To sum up, the AR discovery among discrete ROIs obtained from functional images can simplify the process of learning improving the generalization of the learned results for the early diagnosis of AD.

AR-based FS technique

AR-based FS technique was proposed as an effective way to solve the sample size problem of a functional database (SPECT or PET) for the early diagnosis of AD in order to design a CAD system. Firstly, ARs were obtained (in the same way as the AR baseline) from activated blocks of controls at different *minsup* and *minconf* values. When the maxima values are tuned, that is 100%, the most important voxels are selected and the rest discarded. The voxels selected are formulated in terms of the activation of antecedents and consequents of ARs merged without repetition. The FS dimension is in this case reduced from 2445 to 250 (for SPECT) and from 3776 to 230 (for PET) features. In the following, a subsequent reduction by means of PCA and PLS techniques is necessary to set up the main feature vectors used as input for a kernel SVM classifier (linear for SPECT and rbf for PET modality). Several experiments were conducted in order to test the reliability of the proposed AR FS-based method in terms of Acc, Spe, Sen converging to an optimum operating point of 91.75%, 95.12%, 89.29% respectively for 12 PLS features (SPECT) and 90%, 90.67%, 89.33% (6 PLS features for PET), outperforming other recently reported techniques.

15.1.2 LMNN-based technique

Several experiments were conducted in order to evaluate the proposed LMNN-based feature extraction algorithms and its benefits as: *i*) linear transformation of the PLS or PCA reduced data, *ii*) feature reduction technique, and *iii*) classifier (with Euclidean, Mahalanobis or Energy-based methodology). LMNN classification using energy-based models and Mahalanobis distances performs better than when the Euclidean distance is considered, which suffers a decrease in the accuracy as the number of features increases. Aiming at further improving the accuracy of the classification, SVM was also compared to LMNN-based classification yielding improved results. Thus, the proposed methods yielded Acc rates of 92.7% for SPECT and 90.11% for PET when

an advanced feature extraction technique consisting of NMSE feature selection, PLS feature reduction and LMNN transformation in combination with linear SVM classification was considered, thus outperforming other recently and consolidated reported methods such as VAF, PCA or GMM. One of the principal advantages of this technique is the robustness and stability of the proposed methods shown in this work as stated in the Experiments of chapter 14. Another property is its generalization ability in the light of the results obtained with an PET database.

It is remarkable that the SPECT and PET databases used in this work, are not pathologically confirmed and consequently may introduce some uncertainty on the subject's labels. Using these labels allows to test the robustness of the different classifiers. This should be also considered when comparing to other methods tested on autopsy confirmed AD patients on which every classifier is expected to improve the performance.

15.2 Future work

Finally, these proposed methods could be used not only for quantitative assessment but for visual assessment in clinical practice as shown in the examples of the previous sections. Since the presented work does not depend on any pathological information about the specific disease, it is also applicable to other types of neuro-degenerative diseases as Parkinson for instance. It may also be applied to other image modalities, like for instance fMRI. Furthermore, recent studies have shown that different biomarkers may provide complementary information and an improvement in classification for the diagnosis of AD [Davatzikos et al., 2010], as for example in [Zhang et al., 2011], where different modalities of biomarkers are combined, i.e., MRI, FDG-PET and CSF for PET image classification.

Bibliography

- Adankon, M. M., Cheriet, M., 2009. Model selection for the ls-svm. application to handwriting recognition. *Pattern Recognition* 42 (12), 3264 – 3270.
- Adler, R., 1981. *The Geometry of random fields*. Wiley, New York.
- Agrawal, R., Imielinski, T., Swami, A., Jun. 1993. Mining association rules between sets of items in large databases. *SIGMOD Rec.* 22 (2), 207–216.
- Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules in large databases. In: Bocca, J. B., Jarke, M., Zaniolo, C. (Eds.), *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*. Morgan Kaufmann, pp. 487–499.
- Alipanahi, B., Biggs, M., Ghodsi, A., 2008. Distance metric learning vs. fisher discriminant analysis. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (2008)*, 599–603.
- Allan, C. L., Sexton, C. E., Welchew, D., Ebmeier, K. P., 2010. Imaging and biomarkers for alzheimer's disease. *Maturitas* 65 (2), 138 – 142.
- Álvarez, I., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., López, M., Puntonet, C. G., Segovia, F., 2009. Alzheimer's diagnosis using eigenbrains and support vector machines. *IET Electronics Letters* 45 (7), 342–343.
- Alzheimer, A., 1907. Über eine eigenartige erkrankung der hirnrinde. *Allg Z Psychiat* 64, 146–148.
- Alzimami, K. S., Sassi, S. A., Alfuraih, A. A., Livieratos, L., Spyrou, N. M., 2011. Investigation of the possibility of improving spatial resolution in spect with the combination of labr3:ce-based detector and 3d-osem reconstruction algorithms. *Nuclear Instruments and Methods in Physics Research Section*

- A: Accelerators, Spectrometers, Detectors and Associated Equipment 633, Supplement 1 (0), S289 – S291.
- Angulo, C., Anguita, D., Gonzalez-Abril, L., Ortega, J., 2008. Support vector machines for interval discriminant analysis. *Neurocomputing* 71 (7,9), 1220 – 1229.
- Antonie, M.-L., Zaićane, O. R., Coman, A., 2001. Application of data mining techniques for medical image classification. In: *In Proc. of Second Intl. Workshop on Multimedia Data Mining (MDM KDD 2001) in conjunction with Seventh ACM SIGKDD*. pp. 94–101.
- Ashburner, J., June 2000. Voxel-based morphometry methods. *NeuroImage* 11 (6), 805–821.
- Ashburner, J., 2009. Computational anatomy with the spm software. *Magnetic Resonance Imaging* 27 (8), 1163 – 1174.
- Ashburner, J., 2012. Spm: A history. *NeuroImage* 62 (2), 791 – 800.
- Ashburner, J., Friston, K. J., 1999. Nonlinear spatial normalization using basis functions. *Human Brain Mapping* 7 (4), 254–66.
- Association, A. P., Association, A. P., on DSM-IV, T. F., Jul. 2000. Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision), 4th Edition. Amer Psychiatric Pub.
- B. Scholkopf, A. Smola, K. R. M., 1998. Non-linear component analysis as a kernel eigenvalue problem. *Neural Computation* 10 (5), 1299–1319.
- Bastien, P., Vinzi, V. E., Tenenhaus, M., 2005. Pls generalised linear regression. *Computational Statistics and Data Analysis* 48 (1), 17 – 46.
- Baumgartner, R., Scarth, G., Teichtmeister, C., Somorjai, R., Moser, E., 1997. Fuzzy clustering of gradient-echo functional mri in the human visual cortex part 1: reproducibility. *J Mag. Res. Imaging* (7), 1094–1101.
- Bengio, Y., Grandvalet, Y., 2004. No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research* 5, 1089–1105.
- Bennett, D. A., Evans, D. A., 1992. Alzheimer’s disease. *Disease-a-Month* 38 (1), 7 – 64.

- Benoit, M., Koulibaly, P. M., Migneco, O., Darcourt, J., Pringuey, D. J., Robert, P. H., 2002. Brain perfusion in alzheimer's disease with and without apathy: a spect study with statistical parametric mapping analysis. *Psychiatry Research: Neuroimaging* 114 (2), 103 – 111.
- Berzal, F., 2002. Art: Un método alternativo para la construcción de árboles de decisión. Ph.D. thesis, Universidad de Granada.
- Berzal, F., Cubero, J.-C., Marín, N., Serrano, J. M., 2001. Tbar: an efficient method for association rule mining in relational databases. *Data and Knowledge Engineering* 37, 47 – 64.
- Bishop, C. M., Oct. 2006. *Pattern recognition and machine learning*, 1st Edition. Springer.
- Bonmati, L. M., Alberich-Bayarri, A., García-Martí, G., Requena, R. S., Castillo, C. P., Sierra, J. C., Herrera, J. M., 2012. Biomarcadores de imagen, imagen cuantitativa y bioingeniería. *Radiología* 54 (3), 269–278.
- Borgelt, C., 2005. An implementation of the fp-growth algorithm. In: *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*. OSDM '05. pp. 1–5.
- Bouras, C., Hof, P. R., Giannakopoulos, P., Michel, J.-P., Morrison, J. H., 1994. Regional Distribution of Neurofibrillary Tangles and Senile Plaques in the Cerebral Cortex of Elderly Patients: A Quantitative Evaluation of a One-Year Autopsy Population from a Geriatric Hospital 4 (2), 138–150.
- Braak, H., Tredici, K. D., 2012. Alzheimer's disease: Pathogenesis and prevention. *Alzheimer's and amp; Dementia* 8 (3), 227 – 233.
- Bradberry, T. J., Metman, L. V., Contreras-Vidal, J. L., van den Munckhof, P., Hosey, L. A., Thompson, J. L., Schulz, G. M., Lenz, F., Pahwa, R., Lyons, K. E., Braun, A. R., 2012. Common and unique responses to dopamine agonist therapy and deep brain stimulation in parkinson's disease: An h215o pet study. *Brain Stimulation* 5 (4), 605 – 615.
- Breiman, L., Oct. 1999. Prediction games and arcing algorithms. *Neural Comput.* 11 (7), 1493–1517.
- Brin, S., Motwani, R., Ullman, J. D., Tsur, S., 1997. Dynamic itemset counting and implication rules for market basket data. In: *Proceedings of the 1997 ACM SIGMOD international conference on Management of data*. SIGMOD '97. pp. 255–264.

- Brugnolo, A., Nobili, F., Barbieri, M., Dessi, B., Ferro, A., Girtler, N., Palumeri, E., Partinico, D., Raiteri, U., Regesta, G., Servetto, G., Tanganelli, P., Uva, V., Mazzei, D., Donadio, S., Carli, F. D., Colazzo, G., Serrati, C., Rodriguez, G., 2009. The factorial structure of the mini mental state examination (mmse) in alzheimer's disease. *Archives of Gerontology and Geriatrics* 49 (1), 180 – 185.
- Bruyant, P. P., 2002. Analytic and iterative reconstruction algorithms in spect. *The Journal of Nuclear Medicine* 43 (10), 1343–1358.
- Bucchianico, A. D., 1996. Combinatorics, computer algebra and wilcoxon-mann-whitney test. *J. Statist. Plann. Inference* 79, 349–364.
- Burges, C., Jun. 1998. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2 (2), 121–167.
- Cabella, B. C., Sturzbecher, M. J., de Araujo, D. B., Neves, U. P., 2009. Generalized relative entropy in functional magnetic resonance imaging. *Physica A: Statistical Mechanics and its Applications* 388 (1), 41 – 50.
- Cappa, A., Sanctis, S. D., Benet, B., 2004. P2-329 the influence of cognition and psychiatric symptoms on activities of daily living and lifestyle in alzheimer's disease. *Neurobiology of Aging* 25, Supplement 2 (0), S326 – S327.
- Cardinali, D. P., 2007. *Neurociencia aplicada: Sus fundamentos*. Editorial Médica Panamericana.
- Caroli, A., Frisonia, G., for the Alzheimer's Disease Neuroimaging Initiative, 2010. The dynamics of alzheimer's disease biomarkers in the alzheimer's disease neuroimaging initiative cohort. *Neurobiology of Aging* (31), 1263–1274.
- Chai, J., Liu, H., Chen, B., Bao, Z., 2010. Large margin nearest local mean classifier. *Signal Processing* 90, 236–248.
- Chatfield, C., Collins, A. J., 1980. *Introduction to multivariate analysis*. Chapman and Hall LTD.
- Chatpatanasiri, R., Korsrilabutr, T., Tangchanachaianan, P., Kijsirikul, B., 2010. A new kernelization framework for mahalanobis distance learning algorithms. *Neurocomputing* 73, 1570–1579.

- Chaves, R., Górriz, J. M., Ramírez, J., Illán, I. A., Salas-Gonzalez, D., Gómez-Río, M., 2011a. Efficient mining of association rules for the early diagnosis of alzheimer's disease. *Physics in Medicine and Biology* 56 (18), 6047.
- Chaves, R., Ramírez, J., Górriz, J., Illán, I., Salas-González, D., for the Alzheimer's Disease Neuroimaging Initiative, 2012a. Fdg and pib biomarker pet analysis for the alzheimer's disease detection using association rules. In: *Proceedings of the ieee nuclear science symposium and medical imaging conference*. pp. 2576–2579.
- Chaves, R., Ramírez, J., Górriz, J., Illán, I., Segovia, F., Olivares, A., 2011b. Effective diagnosis of alzheimer's disease by means of distance metric learning and random forest. In: *Proceedings of the IWINAC 2011*. pp. 56–67.
- Chaves, R., Ramírez, J., Górriz, J., López, M., Illán, I., Salas-Gonzalez, D., Segovia, F., 2009a. Spect image classification based on nmse feature correlation weighting and svm. In: *Proceedings of the ieee nuclear science symposium and medical imaging conference*. pp. 2715–2719.
- Chaves, R., Ramírez, J., Górriz, J., Salas-Gonzalez, D., Illán, I., Segovia, F., Padilla, P., 2010. Effective diagnosis of alzheimer's disease by means of association rules. In: *Proceedings of the hybrid artificial intelligence systems, hais 2010*. pp. 452–459.
- Chaves, R., Ramírez, J., Górriz, J., Salas-Gonzalez, D., López, M., 2011c. Distance metric learning as feature reduction technique for the alzheimer's disease diagnosis. In: *Proceedings of the IWINAC 2011*. pp. 68–79.
- Chaves, R., Ramírez, J., Górriz, J., Salas-Gonzalez, D., López, M., Illán, I., Olivares, A., 2011d. Effective diagnosis of alzheimer's disease by means of distance metric learning. In: *Proceedings of the Sixth international conference on hybrid artificial intelligence systems (hais 11)*. pp. 148–155.
- Chaves, R., Ramírez, J., Górriz, J., Segovia, F., for the Alzheimer's Disease Neuroimaging Initiative, 2012b. Effective voxel selection method over fdg-pet database based on association rules. In: *Proceedings of the 16th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*.
- Chaves, R., Ramírez, J., Górriz, J. M., López, M., Salas-Gonzalez, D., Álvarez, I., Segovia, F., 2009b. Svm-based computer aided diagnosis of the alzheimer's disease using t-test nmse feature selection with feature correlation weighting. *Neuroscience Letters* 461 (1), 293–297.

- Chaves, R., Ramírez, J., Górriz, J., for the Alzheimer's Disease Neuroimaging Initiative, 2013. Integrating discretization and association rule-based classification for alzheimer's disease diagnosis. *Experts Systems with Applications* 40 (5), 1571–1578.
- Chaves, R., Ramírez, J., Górriz, J., Illán, I., 2012c. Functional brain image classification using association rules defined over discriminant regions. *Pattern Recognition Letters* 33 (12), 1666–1672.
- Chaves, R., Ramírez, J., Górriz, J., Illán, I., Río, M. G., for the Alzheimer's Disease Neuroimaging Initiative, 2012d. Effective diagnosis of alzheimer's disease by means of large margin-based methodology. *BMC medical informatics and decision making* 12 (79).
- Chaves, R., Ramírez, J., Górriz, J., Puntonet, C., 2012e. Association rule-based feature selection method for alzheimer's disease diagnosis. *Expert Systems with Applications* 39 (14), 11766 – 11774.
- Chen, R., Herskovits, E. H., 2010. Machine-learning techniques for building a diagnostic model for very mild dementia. *NeuroImage* 52 (1), 234 – 244.
- Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., Rei, L., Squarcia, S., Rodriguez, G., Bellotti, R., Cerello, P., Mitri, I. D., Retico, A., Nobili, F., 2011. Local mri analysis approach in the diagnosis of early and prodromal alzheimer's disease. *NeuroImage* 58 (2), 469 – 480.
- Chmielewski, M. R., Grzymala-Busse, J. W., 1994. Global discretization of continuous attributes as preprocessing for machine learning. *Third International Workshop on Rough Sets and Soft Computing*, 294–301.
- Chopra, S., Hadsell, R., LeCun., Y., 2005. Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-05)* San Diego, CA, 349–356.
- Chornoboy, E. S., Chen, C. J., Miller, M. I., Miller, T. R., Snyder, D. L., 1990. An evaluation of maximum likelihood reconstruction for spect. *IEEE Transactions on Medical Imaging* 9 (1), 99–110.
- Christensen, A., Uzzel, B., 2000. *International Handbook of Neuropsychological Rehabilitation*. Springer, New York.

- Collins, D., Evans, A., Holmes, C., Peters, T., 1995. Automatic 3D segmentation of neuroanatomical structures from MRI. *Information Processing in Medical Imaging*, 130–152.
- Conesa, A. G., 2011. Año de la investigación en alzheimer. *Fisioterapia* 33 (6), 33 – 233.
- Contreras, M., Ceric, F., Torrealba, F., 2008. El lado negativo de las emociones: la adicción a drogas de abuso. *REV NEUROL* 47 (9), 471 – 476.
- Costafreda, S. G., Chu, C., Ashburner, J., Fu, C. H. Y., 07 2009. Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS ONE* 4 (7), e6353.
- Courtney, S. M., Ungerleider, L. G., Keil, K., Haxby, J. V., 1997. Transient and sustained activity in a distributed neural system for human working memory. *Nature* 386, 608–612.
- Cummings, J. L., 2009. Defining and labeling disease-modifying treatments for alzheimer's disease. *Alzheimer's and Dementia* 5 (5), 406 – 418.
- Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., Trojanowski, J. Q., 2010. Prediction of mci to ad conversion, via mri, csf biomarkers, and pattern classification. *Neurobiology of Aging* 32 (12), 2322.e19–27.
- de Jong, S., 1993. Simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* 18 (3), 251 – 263.
- Díaz, C., Haro, F. J. D., 2004. TÉCNICAS DE EXPLORACIÓN EN MEDICINA NUCLEAR. Ciclo Formativo Imagen para el Diagnóstico. Elsevier Masson.
- Domínguez-Gadea, L., 2008. Tomografía por emisión de fotón único. In: Urturbe, F. M., Lago, M. R., Alonso, R. C. (Eds.), *Neuroimagen. Técnicas y procesos cognitivos*. Elsevier España, Barcelona, pp. 127 – 154.
- Dougherty, J., Kohavi, R., Sahami, M., 1995. Supervised and unsupervised discretization of continuous features. *Machine learning: proceedings on the twelfth international conference*, 194–202.
- Drótos, G., Pákási, M., Papp, E., Kálmán, J., 2012. P-448 - the validation of adas-cog: cognitive performance and characteristics of patients with

- alzheimer's disease or depressive pseudodementia. *European Psychiatry* 27, Supplement 1 (0), 1 –.
- Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., Scheltens, P., Aug. 2007. Research criteria for the diagnosis of alzheimer's disease: revising the nincds-adrda criteria. *Lancet neurology* 6 (8), 734–746.
- Duchesnay, E., Cachia, A., Boddaert, N., Chabane, N., Mangin, J.-F., Martinot, J.-L., Zilbovicius, M., 2011. Feature selection and classification of imbalanced datasets. application to PET images of children with autistic spectrum disorders. *Neuroimage* 57 (3), 1003–1014.
- Duda, R., Hart, P., Stork, D., Nov. 2001. *Pattern Classification* (2nd Edition), 2nd Edition. Wiley-Interscience.
- Duin, R. P. W., 2000. Classifiers in almost empty spaces. In: *Proceedings 15th International Conference on Pattern Recognition*. Vol. 2. IEEE, pp. 1–7.
- Erlandsson, K., Thomas, B., Dickson, J., Hutton, B. F., 2011. Partial volume correction in spect reconstruction with osem. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 648, Supplement 1 (0), S85 – S88.
- Ewers, M., Walsh, C., Trojanowski, J. Q., Shaw, L. M., Petersen, R. C., Jr., C. R. J., Feldman, H. H., Bokde, A. L., Alexander, G. E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H., 2012. Prediction of conversion from mild cognitive impairment to alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging* 33 (7), 1203 – 1214.e2.
- Fayyad, U., Irani, K., 1993. Multi-interval discretization of continuous-valued attributes for classification learning. In: *Proceedings of the 13th International Joint Conference on Artificial Intelligence*. pp. 1022–1027.
- Feldman, H. H., 2007. *Atlas of Alzheimer's Disease*. Informa Healthcare, London.
- Findeis, M. A., 2007. The role of amyloid β peptide 42 in alzheimer's disease. *Pharmacology and Therapeutics* 116 (2), 266 – 286.

- Fisher, R. A., 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7 (2), 179 – 188.
- Frackowiak, R. S. J., Ashburner, J. T., Penny, W. D., Zeki, S., December 2003. *Human Brain Function, Second Edition*. Academic Press.
- Friston, K., Frith, C., Frackowiak, R., 1991. Comparing functional (pet) images: the assessment of significant change. *J Cereb Blood Flow Metab* 4 (11), 690–699.
- Friston, K. J., Ashburner, J., Kiebel, S. J., Nichols, T. E., Penny, W. D., 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., Frackowiak, R. S. J., 1994. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* 2 (4), 189–210.
- F.Segovia, Noviembre 2010. Análisis de imágenes funcionales cerebrales mediante modelos de mezcla de gaussianas y mínimos cuadrados parciales para el diagnóstico de alteraciones neurológicas. Ph.D. thesis, Universidad de Granada.
- Fujita, H., Uchiyama, Y., Nakagawa, T., Fukuoka, D., Hatanaka, Y., Hara, T., Lee, G. N., Hayashi, Y., Ikedo, Y., Gao, X., Zhou, X., 2008. Computer-aided diagnosis: The emerging of three cad systems induced by japanese health care needs. *Computer Methods and Programs in Biomedicine* 92 (3), 238 – 248.
- Fung, G., Stoeckel, J., 2007. Svm feature selection for classification of spect images of alzheimer’s disease using spatial information. *Knowl. Inf. Syst.* 11 (2), 243–258.
- Gaitanis, A., Kontaxakis, G., Spyrou, G., Panayiotakis, G., Tzanakos, G., 2010. Pet image reconstruction: A stopping rule for the mlem algorithm based on properties of the updating coefficients. *Computerized Medical Imaging and Graphics* 34 (2), 131 – 141.
- Gauthier, S., 2001. Alzheimer’s disease: current and future therapeutic perspectives. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 25 (1), 73 – 89.

- Geng, X., Liu, T.-Y., Qin, T., Li, H., 2007. Feature selection for ranking. In: 30th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 407–414.
- Gerlach, R. W., Kowalski, B. R., Wold, H. O., 1979. Partial least-squares path modelling with latent variables. *Analytica Chimica Acta* 112 (4), 417 – 421.
- Gheyas, I. A., Smith, L. S., 2010. Feature subset selection in large dimensionality domains. *Pattern Recognition* 43 (1), 5 – 13.
- Giasson, B. I., Ischiropoulos, H., Lee, V. M., Trojanowski, J. Q., 2002. The relationship between oxidative/nitrative stress and pathological inclusions in alzheimer's and parkinson's diseases. *Free Radical Biology and Medicine* 32 (12), 1264 – 1275.
- Gidskehaug, L., Anderssen, E., Alsberg, B. K., 2008. Cross model validation and optimisation of bilinear regression models. *Chemometrics and Intelligent Laboratory Systems* 93, 1–10.
- Gillain, S., Warzee, E., Lekeu, F., Wojtasik, V., Maquet, D., Croisier, J.-L., Salmon, E., Petermans, J., 2009. The value of instrumental gait analysis in elderly healthy, mci or alzheimer's disease subjects and a comparison with other clinical tests used in single and dual-task conditions. *Annals of Physical and Rehabilitation Medicine* 52 (6), 453 – 474.
- Gispert, J. D., Pascau, J., Reig, S., García-Barreno, P., Desco, M., 2003. Mapas de estadísticos paramétricos (spm) en medicina nuclear. *Rev. Esp. Med. Nuclear* 22 (1), 43 – 53.
- Güçlü, U., Güçlütürk, Y., Loo, C. K., 2011. Evaluation of fractal dimension estimation methods for feature extraction in motor imagery based brain computer interface. *Procedia Computer Science* 3 (0), 589 – 594.
- Globerson, A., Roweis, S. T., 2005. Metric learning by collapsing classes. *Advances in Neural Information Processing Systems* 18, 451–458.
- Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R., 2005. Neighbourhood components analysis. *Advances in Neural Information Processing Systems*, Cambridge MA 17, 513–520.
- Górriz, J. M., Lassl, A., Ramírez, J., Salas-Gonzalez, D., Puntonet, C. G., Lang, E. W., 2009. Automatic selection of rois in functional imaging using gaussian mixture models. *Neuroscience letters* 460 (2), 108–111.

- Górriz, J. M., Segovia, F., Ramírez, J., Lassel, A., Salas-Gonzalez, D., 2011. Automatic selection of rois in functional imaging using gaussian mixture models. *Applied Soft Computing* 11 (2), 2376–2382.
- Graham, L., Levin, C. S., Muehllehner, G., 2002. Anger Scintillation Camera, In: *Diagnostic Nuclear Medicine*. Lippincott, Williams & Wilkins, Philadelphia, PA, pp. 31–42.
- Guekht, A. B., Moessler, H., Novak, P. H., Gusev, E. I., 2011. Cerebrolysin in vascular dementia: Improvement of clinical outcome in a randomized, double-blind, placebo-controlled multicenter trial. *Journal of Stroke and Cerebrovascular Diseases* 20 (4), 310 – 318.
- Habert, M.-O., Horn, J.-F., Sarazin, M., Lotterie, J.-A., Puel, M., Onen, F., Zanca, M., Portet, F., Touchon, J., Verny, M., Mahieux, F., Giron, A., Fertil, B., Dubois, B., 2011. Brain perfusion spect with an automated quantitative tool can identify prodromal alzheimer’s disease among patients with mild cognitive impairment. *Neurobiology of Aging* 32 (1), 15 – 23.
- Hampel, H., Goernitz, A., Buerger, K., 2003. Advances in the development of biomarkers for alzheimer’s disease: from csf total tau and a β 1 to 42 proteins to phosphorylated tau protein. *Brain Research Bulletin* 61 (3), 243 – 253, laboratory Diagnosis of Brain Damage and Inflammation with New and Old Markers in Human Cerebrospinal Fluid and Blood.
- Hampel, H., Prvulovic, D., Teipel, S., Jessen, F., Luckhaus, C., Frölich, L., Riepe, M. W., Dodel, R., Leyhe, T., Bertram, L., Hoffmann, W., Faltraco, F., 2011. The future of alzheimer’s disease: The next 10 years. *Progress in Neurobiology* 95 (4), 718 – 728.
- Han, G., Zhao, C., 2010. Auc maximization linear classifier based on active learning and its application. *Neurocomputing* 73 (7), 1272 – 1280.
- Hansen, L. K., 2007. Multivariate strategies in functional magnetic resonance imaging. *Brain and Language* 102 (2), 186 – 191.
- Henley, D. B., Sundell, K. L., Sethuraman, G., Siemers, E. R., 2012. Safety profile of alzheimer’s disease populations in alzheimer’s disease neuroimaging initiative and other 18-month studies. *Alzheimer’s and Dementia* 8 (5), 407 – 416.
- Hipp, J., Güntzer, U., Nakhaeizadeh, G., Jun. 2000. Algorithms for association rule mining – a general survey and comparison. *SIGKDD Explor. Newsl.* 2 (1), 58–64.

- Holman, B. L., Johnson, K. A., Gerada, B., Carvalho, P. A., Satlin, A., 1992. The scintigraphic appearance of alzheimers disease: a prospective study using technetium-99m-hmpao spect. *J Nucl Med* 33 (2), 181–5.
- Honeine, P., Noumir, Z., Richard, C., 2013. Multiclass classification machines with the complexity of a single binary classifier. *Signal Processing* 93 (5), 1013 – 1026.
- Houtsma, M., 1995. Set-oriented mining for association rules in relational databases. In: *In 11th Intl. Conf. Data Engineering*. pp. 25–33.
- Hu, D., Yan, L., Liu, Y., Zhou, Z., Friston, K. J., Tan, C., Wu, D., 2005. Unified spm ica for fmri analysis. *NeuroImage* 25 (3), 746 – 755.
- Hu, Q., Zhu, P., Yang, Y., Yu, D., Jan. 2011. Letters: Large-margin nearest neighbor classifiers via sample weight learning. *Neurocomput.* 74 (4), 656–660.
- Huang, S., Xiao, H., 2011. Research on eeg features extraction based on clique. *Procedia Environmental Sciences* 10, Part B (0), 1333 – 1337.
- Hudson, M., Larkin, R., (such As Expectation Maximization, A., 1994. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag* 13, 601–609.
- I. Goethals, C. Van De Wiele, D. S., Dierckx, R., 2002. Brain spet perfusion in early alzheimer’s disease: where to look? *European Journal of Nuclear Medicine and Molecular Imaging* 29, 975–978.
- I.A.Illán, July 2009. Análisis en componentes de imágenes funcionales para la ayuda al diagnóstico de la enfermedad de alzheimer. Ph.D. thesis, Universidad de Granada.
- Illán, I., Górriz, J., Ramírez, J., Chaves, R., Segovia, F., López, M., Salas-Gonzalez, D., Puntonet, C., 2010. Machine learning for very early alzheimer’s disease diagnosis, a 18f-fdg and pib pet comparison. In: *IEEE Nuclear Science Symposium and medical image conference*. pp. 2334 – 2337.
- Illán, I. A., Górriz, J. M., López, M. M., Ramírez, J., Salas-Gonzalez, D., Segovia, F., Chaves, R., Puntonet, C. G., Mar. 2011a. Computer aided diagnosis of alzheimer’s disease using component based svm. *Appl. Soft Comput.* 11 (2), 2376–2382.

- Illán, I. A., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., López, M., Segovia, F., Padilla, P., Puntonet, C. G., Aug. 2010. Projecting independent components of spect images for computer aided diagnosis of alzheimer's disease. *Pattern Recogn. Lett.* 31 (11), 1342–1347.
- Illán, I. A., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., López, M. M., Segovia, F., Chaves, R., Gómez-Río, M., Puntonet, C. G., Feb. 2011b. 18f-fdg pet imaging analysis for computer aided alzheimer's diagnosis. *Inf. Sci.* 181 (4), 903–916.
- Illán, I. A., Górriz, J. M., Ramírez, J., Chaves, R., Segovia, F., Lóèz, M., Salas-Gonzalez, D., Puntonet, C. G., 2010. Machine learning for very early ad diagnosis;a 18f-fdg and pib pet comparison. In: *IEEE Nuclear Science Symposium and Medical Image Conference*. pp. 2334 – 2337.
- Ishii, K., Sasaki, A. K., Miyamoto, N., Fakuda, T., Sakamoto, S., Mori, E., 2006. Fully automatic diagnostic system for early- and late-onset mild alzheimer's disease using fdg pet and 3d-ssp. *European Journal of Nuclear Medicine and Molecular Imaging* 33 (5), 575–583.
- Ito, K., 2006. Pet spect for dementia early diagnosis of alzheimer's disease. *International Congress Series* 1290 (0), 123 – 127.
- Jack, C., Albert, M., Knopman, D., McKhann, G., Sperling, R., Carrillo, M., Thies, B., Phelps, C., 2011. Introduction to the recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's and Dementia* 7 (3), 257–262.
- Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., Trojanowski, J. Q., Jan. 2010. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *Lancet neurology* 9 (1), 119–128.
- Jackson, J. E. (Ed.), 1991. *A Users Guide to Principal Components*. Wiley, New York.
- Jobst, K. A., Barnetson, L. P., Shepstone, B. J., Sep. 1998. Accurate prediction of histologically confirmed alzheimer's disease and the differential diagnosis of dementia: the use of nincds-adrda and dsm-iii-r criteria, spect, x-ray ct, and apo e4 in medial temporal lobe dementias. oxford project to investigate memory and aging. *Int Psychogeriatr* 10 (3), 271–302.
- Jolliffe, I. T., Oct. 2002. *Principal Component Analysis*, 2nd Edition. Springer.

- Karabatak, M., Ince, M. C., Dec. 2009. A new feature selection method based on association rules for diagnosis of erythematous-squamous diseases. *Expert Syst. Appl.* 36 (10), 12500–12505.
- Kettaneh, N., Berglund, A., Wold, S., 2005. Pca and pls with very large data sets. *Computational Statistics and Data Analysis* 48 (1), 69 – 85.
- Khachaturian, Z. S., 2007. A chapter in the development of alzheimer’s disease research: A case study of public policies on the development and funding of research programs. *Alzheimer’s and Dementia* 3 (3), 243 – 258.
- Klunk, W. E., et al., 2004. Imaging brain amyloid in alzheimer’s disease with pittsburgh compound-b. *Annals of Neurology* 55 (3), 306–19.
- Kogure, D., Matsuda, H., Ohnishi, T., Asada, T., Uno, M., Kunihiro, T., Nakano, S., Takasaki, M., 2000. Longitudinal evaluation of early alzheimer’s disease using brain perfusion spect. *J Nucl Med* 41 (7), 1155–1162.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence (IJCAI)*.
- Kohavi, R., John, G. H., 1997. Wrappers for feature subset selection. *ARTIFICIAL INTELLIGENCE* 97 (1), 273–324.
- Kubota, K., Watanabe, H., Murata, Y., Yukihiro, M., Ito, K., Morooka, M., Minamimoto, R., Hori, A., Shibuya, H., 2011. Effects of blood glucose level on fdg uptake by liver: a fdg-pet/ct study. *Nuclear Medicine and Biology* 38 (3), 347 – 351.
- Kuo, R., Chao, C., Chiu, Y., 2011. Application of particle swarm optimization to association rule mining. *Applied Soft Computing* 11 (1), 326 – 336.
- Li, W., Han, J., Pei, J., 2001. Cmar: Accurate and efficient classification based on multiple class-association rules. *Data Mining, IEEE International Conference on*, 369.
- Li, Y., Lu, B.-L., Sep. 2009. Feature selection based on loss-margin of nearest neighbor classification. *Pattern Recogn.* 42 (9), 1914–1921.
- Lillo, P., Hodges, J. R., 2009. Frontotemporal dementia and motor neurone disease: Overlapping clinic-pathological disorders. *Journal of Clinical Neuroscience* 16 (9), 1131 – 1135.

- Liu, B., Hsu, W., Ma, Y., 1998a. Discretization Techniques: A recent survey. *Knowledge Discovery and Data Mining*, 80–86.
- Liu, B., Hsu, W., Ma, Y., 1998b. Integrating classification and association rule mining. In: *Knowledge Discovery and Data Mining*. pp. 80–86.
- Liu, Z., de Zwart, J. A., van Gelderen, P., Kuo, L.-W., Duyn, J. H., 2012. Statistical feature extraction for artifact removal from concurrent fmri-eeeg recordings. *NeuroImage* 59 (3), 2073 – 2087.
- López, M., Ramírez, J., Górriz, J. M., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R., 2009a. Svm-based cad system for early detection of the alzheimer’s disease using kernel pca and lda. *Neuroscience Letters* 464 (4), 233–238.
- López, M., Ramírez, J., Górriz, J. M., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R., Padilla, P., Gómez-Río, M., Mar. 2011. Principal component analysis-based techniques and supervised classification schemes for the early detection of alzheimer’s disease. *Neurocomput.* 74 (8), 1260–1271.
- López, M., Ramírez, J., Górriz, J. M., Salas-Gonzalez, D., Álvarez, I., Segovia, F., Puntonet, C. G., 2009b. Automatic tool for the alzheimers disease diagnosis using pca and bayesian classification rules. *IET Electronics Letters* 45 (8), 389–391.
- Lukic, A. S., Wernick, M. N., Strother, S. C., 2002. An evaluation of methods for detecting brain activations from functional neuroimages. *Artificial Intelligence in Medicine* 25 (1), 69 – 88.
- Maldonado, S., Weber, R., Basak, J., 2011. Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences* 181, 115–128.
- Martí-Climent, J., Prieto, E., Lafuente, J. L., Arbizu, J., 2010. Neuroimagen: Fundamentos técnicos y prácticos. *Revista Española de Medicina Nuclear* 29 (4), 189 – 210.
- Martin, L., Latypova, X., Terro, F., 2011. Post-translational modifications of tau protein: Implications for alzheimer’s disease. *Neurochemistry International* 58 (4), 458 – 471.
- Martínez, F. J., Salas-González, D., Górriz, J. M., Ramírez, J., Puntonet, C. G., Gómez-Río, M., 2011. Analysis of spect brain images using wilcoxon

- and relative entropy criteria and quadratic multivariate classifiers for the diagnosis of alzheimer's disease. In: Proceedings of the 4th international conference on Interplay between natural and artificial computation: new challenges on bioinspired applications - Volume Part II. IWINAC'11. pp. 41–48.
- Martínez-Murcia, F. J., Górriz, J. M., Ramírez, J., Puntonet, C. G., Salas-González, D., Aug. 2012. Computer aided diagnosis tool for alzheimer's disease based on mann-whitney-wilcoxon u-test. *Expert Syst. Appl.* 39 (10), 9676–9685.
- Marui, W., Iseki, E., Nakai, T., Miura, S., Kato, M., Uéda, K., Kosaka, K., 2002. Progression and staging of lewy pathology in brains from patients with dementia with lewy bodies. *Journal of the Neurological Sciences* 195 (2), 153 – 159.
- Mason, L., Baxter, J., Bartlett, P., Frean, M., 2000. Boosting algorithms as gradient descent. In: *In Advances in Neural Information Processing Systems* 12. MIT Press, pp. 512–518.
- Matsuda, H., Mizumura, S., Nagao, T., Ota, T., Iizuka, T., Nemoto, K., Takemura, N., Arai, H., Homma, A., 2007. Automated discrimination between very early alzheimer disease and controls using an easy z-score imaging system for multicenter brain perfusion single-photon emission tomography. *AJNR Am J Neuroradiol* 28 (4), 731–6.
- McKeown, M. J., Jung, T. P., Makeig, S., Brown, G., Kindermann, S. S., Lee, T. W., Sejnowski, T. J., 1998. Spatially independent activity patterns in functional mri data during the stroop color-naming task. *Proc Natl Acad Sci U S A* 95 (3), 803–810.
- Meda, S. A., Narayanan, B., Liu, J., Perrone-Bizzozero, N. I., Stevens, M. C., Calhoun, V. D., Glahn, D. C., Shen, L., Risacher, S. L., Saykin, A. J., Pearlson, G. D., 2012. A large scale multivariate parallel ica method reveals novel imaging genetic relationships for alzheimer's disease in the adni cohort. *NeuroImage* 60 (3), 1608 – 1621.
- Menzies, T., Hu, Y., Nov. 2003. Data mining for very busy people. *Computer* 36 (11), 22–29.
- Metz, C. E., Oct. 1978. Basic principles of roc analysis. *Seminars in nuclear medicine* 8 (4), 283–298.

- Mika, S., Rätsch, G., Weston, J., Schölkopf, B., Müller, K.-R., 1999. Fisher discriminant analysis with kernels.
- Minoshima, S., Frey, K., Koeppe, R., Foster, N., Kuhl, D., 1995. A diagnostic approach in alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-fdg pet. *J Nucl Med* 36 (7), 1238–48.
- Modinos, G., Pettersson-Yeo, W., Allen, P., McGuire, P. K., Aleman, A., Mechelli, A., 2012. Multivariate pattern classification reveals differential brain activation during emotional processing in individuals with psychosis proneness. *NeuroImage* 59 (3), 3033 – 3041.
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., Trojanowski, J. Q., Toga, A. W., Beckett, L., Jul. 2005. Ways toward an early diagnosis in alzheimer's disease: The alzheimer's disease neuroimaging initiative (adni). *Alzheimers Dement* 1 (1), 55–66.
- Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B., 2001. An introduction to kernel-based learning algorithms. *IEEE TRANSACTIONS ON NEURAL NETWORKS* 12 (2), 181–201.
- Nadeau, C., Bengio, Y., 2003. Inference for the generalization error. *Machine Learning* 52 (3), 239–281.
- Nakano, S., Kato, T., Nakamura, S., Kameyama, M., 1986. Acetylcholinesterase activity in cerebrospinal fluid of patients with alzheimer's disease and senile dementia. *Journal of the Neurological Sciences* 75 (2), 213 – 223.
- Nakib, A., Oulhadj, H., Siarry, P., 2007. Image histogram thresholding based on multiobjective optimization. *Signal Processing* 87 (11), 2516 – 2534.
- Nations, U., 2008. Department of economic and social affairs. world population prospects. the 2008 revision.
- Newman, J., von Cramon, D. Y., Lohmann, G., 2008. Model-based clustering of meta-analytic functional imaging data. *NeuroImage* 29, 177–192.
- Ng, S., Villemagne, V. L., Berlangieri, S., Lee, S.-T., Cherk, M., Gong, S. J., Ackermann, U., Saunderson, T., Tochon-Danguy, H., Jones, G., Smith, C., O, Keefe, G., Masters, C. L., Rowe, C. C., 2007. Visual assessment versus quantitative assessment of 11c-pib pet and 18f-fdg pet for detection of alzheimer's disease. *J Nucl Med* 48 (4), 547–52.

- Nitrini, R., Buchpiguel, C., Caramelli, P., Bahia, V., Mathias, S., Nascimento, C., Degenszajn, J., Caixeta, L., 2000. Spect in alzheimer's disease: features associated with bilateral parietotemporal hypoperfusion. *Acta Neurol Scand* 101 (3), 172–6.
- Nofal, A. A. D. M., Ahmad, S. B., 2010. Classification based on association-rule mining techniques: a general survey and empirical comparative evaluation. *UbiCC* 5 (3).
- Noori, R., Karbassi, A., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M., Farokhnia, A., Gousheh, M. G., 2011. Assessment of input variables determination on the svm model performance using pca, gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology* 401 (3), 177–189.
- Ordoñez, C., Omiecinski, E., 1998. Image mining: a new approach for data mining. Tech. rep.
- Ordoñez, C., Santana, C., de Braal, L., 2000. Discovering interesting association rules in medical data. In: *Proceedings of ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*. pp. 78–85.
- Ostwald, D., Bagshaw, A. P., 2011. Information theoretic approaches to functional neuroimaging. *Magnetic Resonance Imaging* 29 (10), 1417 – 1428.
- Padilla, P., Górriz, J., Ramírez, J., Chaves, R., Segovia, F., Illán, I., Salas-Gonzalez, D., López, M., Puntonet, C., 2010. Alzheimer's disease detection in functional images using 2d gabor wavelet analysis. *Electronics Letters* 46 (8), 556–558.
- Park, C. H., Lee, M., Apr. 2009. A svm-based discretization method with application to associative classification. *Expert Syst. Appl.* 36 (3), 4784–4787.
- Pérez, P., Chardin, A., Laferte, J., 2000. Noniterative manipulation of discrete energy-based models for image analysis. *Pattern Recognition* 33, pp. 573–586.
- Pfefferbaum, A., Sullivan, E. V., Jernigan, T. L., Zipursky, R. B., Rosenbloom, M. J., Yesavage, J. A., Tinklenberg, J. R., 1990. A quantitative analysis of ct and cognitive measures in normal aging and alzheimer's disease. *Psychiatry Research: Neuroimaging* 35 (2), 115 – 136.

- Polat, K., Güneş, S., Sep. 2009. A new feature selection method on classification of medical datasets: Kernel f-score feature selection. *Expert Syst. Appl.* 36 (7), 10367–10373.
- Pérez, M. L., July 2010. Nuevos modelos estadísticos para detección de patrones de hipo/perfusión-metabolismo en imágenes de tomografía funcional cerebral. Ph.D. thesis, Universidad de Granada.
- Pupi, A., Cristofaro, M. D., Nacmias, B., Sorbi, S., Mosconi, L., 2009. Brain glucose metabolism: Age, alzheimer's disease, and apoe allele effects. In: in Chief: Larry R. Squire, E. (Ed.), *Encyclopedia of Neuroscience*. Academic Press, Oxford, pp. 363 – 373.
- qiao Peng, M. J., Ju, X., Khambay, B. S., Ayoub, A. F., Chen, C.-T., Bai, B., 2012. Clinical significance of creative 3d-image fusion across multimodalities [pet;ct;mr] based on characteristic coregistration. *European Journal of Radiology* 81 (3), e406 – e413.
- Ramírez, J., Górriz, J. M., Chaves, R., Salas-Gonzalez, D., Álvarez, I., López, M., Segovia, F., 2009a. Spect image classification using random forests. *IET Electronics Letters* 45 (12), 604–605.
- Ramírez, J., Górriz, J. M., Salas-Gonzalez, D., López, M., Álvarez, I., Gómez-Río, M., 2009b. Computer aided diagnosis of alzheimer type dementia combining support vector machines and discriminant set of features. *Information Sciences* In press.
- Ramírez, J., Górriz, J. M., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., Álvarez, I., Padilla, P., 2010. Early alzheimer's disease diagnosis using partial least squares and random forests. In: *Proceedings of the 2010 IEEE international conference on Biomedical imaging: from nano to Macro. ISBI'10*. IEEE Press, Piscataway, NJ, USA, pp. 81–84.
- Refaeilzadeh, P., Tang, L., Lui, H., 2008. k-fold cross-validation.
- Reilhac, A., Tomeï, S., Buvat, I., Michel, C., Keheren, F., Costes, N., 2008. Simulation-based evaluation of osem iterative reconstruction methods in dynamic brain pet studies. *NeuroImage* 39 (1), 359 – 368.
- Reinikainen, K., Pitkänen, A., Riekkinen, P., 1989. 2',3'-cyclic nucleotide-3'-phosphodiesterase activity as an index of myelin in the post-mortem brains of patients with alzheimer's disease. *Neuroscience Letters* 106 (1), 229 – 232.

- Ribeiro, M. X., Bugatti, P. H., Traina, Jr., C., Marques, P. M. A., Rosa, N. A., Traina, A. J. M., Dec. 2009. Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques. *Data Knowl. Eng.* 68 (12), 1370–1382.
- Ribeiro, M. X., Marques, J., Traina, A. J. M., Traina Jr., C., 2006. Statistical association rules and relevance feedback: Powerful allies to improve the retrieval of medical images. In: *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems. CBMS '06.* IEEE Computer Society, Washington, DC, USA, pp. 887–892.
- Rosa-Neto, P., Leuzy, A., 2009. Molecular imaging of alzheimer's disease using pet. *The Canadian Review of Alzheimer's Disease and Other Dementias*, 18–24.
- Rosipal, R., 2003. Kernel partial least squares for nonlinear regression and discrimination. *Neural Network World* 13(3), 291–300.
- Rubio, J. C., de Bustos, F., Molina, J. A., Jiménez-Jiménez, F. J., Benito-León, J., Martín, M. A., Campos, Y., Orti'-Pareja, M., Cabrera-Valdivia, F., Arenas, J., 1998. Cerebrospinal fluid carnitine levels in patients with alzheimer's disease. *Journal of the Neurological Sciences* 155 (2), 192 – 195.
- S. Kotsiantis, D. K., 2006. Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering* 32 (1), 47–58.
- S. Lemm, B. Blankertz, T. D. K. M., 2011. Introduction to machine learning for brain imaging. *NeuroImage* 56 (2), 387–399.
- Salas-Gonzalez, D., Diciembre 2009. Nuevos algoritmos de registro y clasificación de imágenes cerebrales pet y spect para el diagnóstico de la enfermedad de alzheimer. Ph.D. thesis, Universidad de Cadiz.
- Salas-Gonzalez, D., Górriz, J., Ramírez, J., Illán, I., López, M., Segovia, F., Chaves, R., Puntonet, C., 2010a. Feature selection using factor analysis for alzheimer's diagnosis using f-fdg pet images. *Medical Physics*, 6084–6095.
- Salas-Gonzalez, D., Górriz, J., Ramírez, J., López, M., Illán, I., Chaves, R., Segovia, F., Puntonet, C., 2010b. Computer-aided diagnosis of alzheimer's disease using support vector machines and classification trees. *Physics in Medicine and Biology* 55 (10), 2807–2814.

- Salas-Gonzalez, D., Górriz, J. M., Ramírez, J., Lassl, A., Puntonet, C. G., October 2008. Improved gauss-newton optimization methods in affine registration of spect brain images. *IET Electronics Letters* 44 (22), 1291–1292.
- Saxena, P., Pavel, D. G., Quintana, J. C., Horwitz, B., 1998. An automatic threshold-based scaling method for enhancing the usefulness of tc-hmpao spect in the diagnosis of alzheimer's disease. In: *MICCAI*. pp. 623–630.
- Schechtman, E., Sherman, M., 2007. The two-sample t -test with a known ratio of variances. *Statistical Methodology* 4 (4), 508 – 514.
- Schellenberg, G., Bird, T., Wijsman, E., Moore, D., Martin, G., 1989. The genetics of alzheimer's disease. *Biomedicine and amp; Pharmacotherapy* 43 (7), 463 – 468.
- Segovia, F., Górriz, J., Ramírez, J., Salas-González, D., Álvarez, I., López, M., Chaves, R., Padilla, P., 2010. Classification of functional brain images using a gmm-based multi-variate approach. *Neuroscience Letters* 474 (1), 58 – 62.
- Sen, A., Biswas, G., 1985. Decision support systems: An expert systems approach. *Decision Support Systems* 1 (3), 197 – 204.
- Sharma, A., Paliwal, K. K., 2008. Cancer classification by gradient lda technique using microarray gene expression data. *Data and amp; Knowledge Engineering* 66 (2), 338 – 347.
- Sheela, L., Shanthi, V., may 2009. An image analysis and classification protocol for characterization of normal and abnormal memory loss in aging from structural mri. In: *2009 International Conference on Signal Processing Systems*. pp. 639 –643.
- Shepp, L. A., Vardi, Y., 1982. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging* 1 (2), 113–122.
- Shin, J., Lee, S.-Y., Kim, S. J., Kim, S.-H., Cho, S.-J., Kim, Y.-B., 2010. Voxel-based analysis of alzheimer's disease pet imaging using a triplet of radiotracers: Pib, fddnp, and fdg. *NeuroImage* 52 (2), 488 – 496.
- Silverman, D. H., Small, G. W., Chang, C. Y., Lu, C. S., Aburto, M. A. K. D., Chen, W., Czernin, J., Rapoport, S. I., Pietrini, P., Alexander, G. E., Schapiro, M. B., Jagust, W. J., Hoffman, J. M., Welsh-Bohmer, K. A., Alavi, A., Clark, C. M., Salmon, E., de Leon, M. J., Mielke, R., Cummings,

- J. L., Kowell, A. P., Gambhir, S. S., Hoh, C. K., Phelps, M. E., Nov. 2001. Positron emission tomography in evaluation of dementia: Regional brain metabolism and long-term outcome. *JAMA: The Journal of the American Medical Association* 286 (17), 2120–2127, PMID: 11694153.
- Slosman, D. O., Ludwig, C., Zerarka, S., Pellerin, L., Chicherio, C., de Ribaupierre, A., Annoni, J.-M., Bouras, C., Herrmann, F., Michel, J.-P., Giacobini, E., Magistretti, P. J., 2001. Brain energy metabolism in alzheimer's disease: 99mtc-hmpao spect imaging during verbal fluency and role of astrocytes in the cellular mechanism of 99mtc-hmpao retention. *Brain Research Reviews* 36 (2), 230 – 240.
- Sánchez, I. P., 2001. Pet radiopharmaceuticals. *Rev Esp Med Nucl* 20 (6), 477–98.
- Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jr., C. R. J., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., Wagster, M. V., Phelps, C. H., 2011. Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's and Dementia* 7 (3), 280 – 292.
- Srikant, R., Agrawal, R., 1995. Mining generalized association rules. In: *Proceedings of the 21th International Conference on Very Large Data Bases. VLDB '95*. pp. 407–419.
- Stoeckel, J., Marzo 2003. Outils de classification pour l'aide au diagnostic: application à la maladie d'alzheimer et à d'autres pathologies cérébrales. Ph.D. thesis, L'Ecole Nationale Supérieure des Mines de Paris.
- Stoeckel, J., Ayache, N., Malandain, G., Koulibaly, P. M., Ebmeier, K. P., Darcourt, J., 2004. Automatic classification of SPECT images of alzheimer's disease patients and control subjects. *Lecture Notes in Computer Science* 3217, 654–662.
- Stoeckel, J., Malandain, G., Migneco, O., Koulibaly, P. M., Robert, P., Ayache, N., Darcourt, J., 2001. Classification of SPECT images of normal subjects versus images of alzheimer's disease patients. *Lecture Notes in Computer Science* 2208, 666–674.

- Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2), 111–147.
- Summa, M. D., Grosso, A., Locatelli, M., 2011. Complexity of the critical node problem over trees. *Computers & Operations Research* 38 (12), 1766 – 1774.
- Swerdlow, R., 2011. Brain aging, alzheimer’s disease, and mitochondria. *Biochim Biophys Acta* 1812(12), 1630–1639.
- Symms, M., Jäger, H. R., Schmierer, K., Yousry, T. A., 2004. A review of structural magnetic resonance neuroimaging. *J Neurol Neurosurg Psychiatry* 75, 1235–1244.
- Talairach, J., Tournoux, P., Jan. 1988. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging (Thieme Classics)*. Thieme.
- Tan, P.-N., Steinbach, M., Kumar, V., 2005. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Tarantola, G., Zito, F., Gerundini, P., 2003. Pet instrumentation and reconstruction algorithms in whole-body applications. *Journal of Nuclear Medicine* 44 (5), 756–769.
- Tenenhaus, M. (Ed.), 1998. *La Regression PLS: Theorie et Pratique*. Technip, Paris.
- Thabtah, F., Cowling, P., Peng, Y., 2004. Mmac: A new multi-class, multi-label associative classification approach. In: *Proceedings of the 4th IEEE International Conference on Data Mining, ICDM '04*. pp. 217–224.
- Thal, D. R., Arendt, T., Waldmann, G., Holzer, M., Zedlick, D., Rüb, U., Schober, R., 1998. Progression of neurofibrillary changes and pht-t in end-stage alzheimer’s disease is different from plaque and cortical microglial pathology. *Neurobiology of Aging* 19 (6), 517 – 525.
- Toga, W., 1999. *Brain Warping*. Academic Press.
- Toivonen, H., 1996. Sampling large databases for association rules. In: *Proceedings of the 22nd Int’l Conference on very Large Databases*. pp. 134–145.

- Tondelli, M., Wilcock, G. K., Nichelli, P., Jager, C. A. D., Jenkinson, M., Zamboni, G., 2012. Structural mri changes detectable up to ten years before clinical alzheimer's disease. *Neurobiology of Aging* 33 (4), 825.e25 – 825.e36.
- Tsochantaridis, I., Joachims, T., Hofmann, T., Altun, Y., 2005. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research* (6), 1453–1484.
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., Zeffiro, T. A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage* 16, 765–780–.
- Umarani, V., Punithavalli, D. M., 2010. Sampling based association rules mining a recent overview. (IJCSE) *International Journal on Computer Science and Engineering* 2, 314–318.
- Vapnik, V. N., 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin.
- Vapnik, V. N., 1998. *Statistical Learning Theory*. John Wiley and Sons Inc., New York.
- Varmuza, K., Filzmoser, P. (Eds.), 2009. *Introduction to Multivariate Statistical Analysis in Chemometrics*. Taylor and Francis - CRC Press, Boca Raton, FL.
- Vellas, B., Froelich, L., Sampaio, C., 2007. Commentary on health economics and the value of therapy in alzheimer's disease. value therapy for alzheimer's disease a european perspective. *Alzheimer's and Dementia* 3, 152–156.
- Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., Boeve, B. F., Petersen, R. C., Jr., C. R. J., 2008. Alzheimer's disease diagnosis in individual subjects using structural mr images: Validation studies. *NeuroImage* 39 (3), 1186 – 1197.
- Villemagne, V. L., Rowe, C. C., Macfarlane, S., Novakovic, K., Masters, C., Apr. 2005. *Imagem oblivionis: the prospects of neuroimaging for early detection of alzheimer's disease*. *J Clin Neurosci* 12 (3), 221–230.
- Wang, L., Xie, K., Chen, T., Ma, X., 2005. Efficient discovery of multilevel spatial association rules using partitions. *Information and Software Technology* 47 (13), 829 – 840.

- Wang, X.-Y., Chen, J.-W., Yang, H.-Y., 2011. A new integrated svm classifiers for relevance feedback content-based image retrieval using em parameter estimation. *Applied Soft Computing* 11 (2), 2787 – 2804.
- Webb, A., 2002. *Statistical Pattern Recognition*, 2nd Edition. John Wiley & Sons.
- Wechalekar, K., Cook, G. J., 2005. Perioperative use of radioisotopes. *Surgery (Oxford)* 23 (5), 180 – 184.
- Weinberger, K., Saul, L. K., 2008. Fast solvers and efficient implementations for distance metric learning. *Proceedings of the 25th international conference on Machine Learning*.
- Weinberger, K. Q., Blitzer, J., Saul, L. K., 2009. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research* 10, 207–244.
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Siuciak, J. A., Soares, H., Toga, A. W., Trojanowski, J. Q., 2012. The alzheimers disease neuroimaging initiative: A review of papers published since its inception. *Alzheimer's and Dementia* 8 (1, Supplement), S1 – S68.
- Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J. S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L., 2011. Multivariate analysis of mri data for alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage* 54 (2), 1178–1187.
- Wiens, T. S., Dale, B. C., Boyce, M. S., Kershaw, G. P., 2008. Three way k-fold cross-validation of resource selection functions. *Original Research Article Ecological Modelling* 212 (3-4), 244–255.
- Wold, H., 1982. Soft modeling. the basic design and some extensions. *Joreskog, K.-G., Wold, H. (Eds.), Systems Under Indirect Observation* 2.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal components analysis. *Chemometr. Intell. Lab. System* 2, 37–52.
- Wolfe, M. S., 2002. App, notch, and presenilin: molecular pieces in the puzzle of alzheimer's disease. *International Immunopharmacology* 2 (13), 1919 – 1929.

- Woods, R. P., Cherry, S. R., Mazziotta, J. C., 1992. Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography* 16 (4), 620–633.
- Wu, H.-C., 2009. The karush kuhn tucker optimality conditions in multiobjective programming problems with interval valued objective functions. *European Journal of Operational Research* 196 (1), 49 – 60.
- Wu, X., Zhang, C., Zhang, S., Jul. 2004. Efficient mining of both positive and negative association rules. *ACM Trans. Inf. Syst.* 22 (3), 381–405.
- Wu, Y., Ianakiev, K., Govindaraju, V., 2002. Improved k-nearest neighbor classification. *Pattern Recognition* 35 (10), 2311 – 2318.
- Xiang, S., Nie, F., Zhang, C., 2008. Learning a mahalanobis distance metric for data clustering and classification. *Pattern Recognition* 41, pp. 3600–3612.
- yan Qu, H., Xu, F., fang Hu, X., bin Wang, L., Zhao, J., Zhang, Z., 2012. A novel denoising method based on radon transform and filtered back-projection reconstruction algorithm. *Optics and Lasers in Engineering* 50 (4), 593 – 598.
- Yang, L., Jin, R., 2006. Distance metric learning: a comprehensive survey. Michigan State University.
- Yin, X., Han, J., 2003. Cpar: Classification based on predictive association rules.
- Yoon, C., Lee, W., Lee, T., 2011. Simulation for czt compton pet (maximization of the efficiency for pet using compton event). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 652 (1), 713 – 716.
- Zaiane, O. R., Antonie, M. L., Coman, A., 2002. Mammography classification by an association rule-based classifier.
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., 2011. Multimodal classification of alzheimer’s disease and mild cognitive impairment. *NeuroImage* 55 (3), 856–867.
- Zubi, Z. S., Saad, R. A., 2011. Using some data mining techniques for early diagnosis of lung cancer. In: *Proceedings of the 10th WSEAS international conference on Artificial intelligence, knowledge engineering and data bases. AIKED’11.* World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, pp. 32–37.