



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Anàlisi d'imatge mèdica del cervell per donar suport al diagnòstic

TREBALL FI DE GRAU

Grau en Enginyeria Informàtica

Autor: Sílvia Nadal Almela

Tutor: Jon Ander Gómez Adrián
Maria de la Iglesia Vayá

Curs 2018-2019

A mon iaio

Agraïments

Vull donar les gràcies a Jon Ander per confiar en mi i haver accedit a tutoritzar aquest projecte. A Álvaro i a Rafa per ser uns companys de confiança i saber treballar en comú. A Mariam i a Paco per guiar-nos durant la nostra estada en el CIPF i ajudar-nos, sobretot, a millorar les nostres habilitats de comunicació. A Jose per tota l'ajuda prestada. A Quim i a Jhon per resoldre'ns els xicotets dubtes que ens sorgien. I també a NVIDIA Corporation per la donació de la GPU NVIDIA Quadro P5000 utilitzada en aquest projecte.

Resum

Durant aquest projecte s'ha treballat amb imatges mèdiques disponibles gràcies a la col·laboració entre el centre d'investigació Pattern Recognition and Human Language Technology –PRHLT– i la Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana –FISABIO–. La finalitat última d'aquest projecte ha sigut construir un classificador per a cada tipus d'imatge segons l'etapa de la malaltia Alzheimer en què es troba el pacient. Els classificadors han estat basats tant en Deep Learning, alguns utilitzant Convolutional Neural Networks, com en mètodes clàssics de Machine Learning. Abans d'entrenar els classificadors, s'ha avaluat la qualitat dels conjunts de dades i s'ha fet un estudi de les seues característiques. Finalment, s'han comparat els resultats obtinguts amb les diferents tècniques emprades.

Paraules clau: xarxes neuronals, imatge mèdica, processament d'imatge, classificador, aprenentatge automàtic

Resumen

Durante este proyecto se ha trabajado con imágenes médicas disponibles gracias a la colaboración entre el centro de investigación Pattern Recognition and Human Language Technology –PRHLT– y la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana –FISABIO–. La finalidad última de este proyecto ha sido construir un clasificador para cada tipo de imagen según la etapa de la enfermedad Alzheimer en que se encuentra el paciente. Los clasificadores se han basado tanto en Deep Learning, algunos utilizando Convolutional Neural Networks, como en métodos clásicos de Machine Learning. Antes de entrenar los clasificadores, se ha evaluado la calidad de los conjuntos de datos y se ha hecho un estudio de sus características. Finalmente, se han comparado los resultados obtenidos con las diferentes técnicas empleadas.

Palabras clave: redes neuronales, imagen médica, procesamiento de imagen, clasificador, aprendizaje automático

Abstract

During this project we have worked with medical images that are available thanks to the collaboration between the research center Pattern Recognition and Human Language Technology (PRHLT) and the Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO). The final goal of this project has been building a classifier for each kind of image according to the stage of Alzheimer's disease of each patient. These classifiers have been based on both Deep Learning, some using Convolutional Neural Networks, and classical methods of Machine Learning. Before training the classifiers, the quality of the data sets has been evaluated and a study of their characteristics has been performed. Finally, the different results obtained with these techniques have been compared.

Key words: neural networks, medical image, image processing, classifier, machine learning

Índex

Índex	ix
Índex de figures	xi
Índex de taules	xii
1 Introducció	1
1.1 Motivació i objectius	1
1.2 Estructura de la memòria	2
1.3 Col·laboracions	2
2 Estat de l'Art	3
2.1 Crítica a l'Estat de l'Art	4
3 Descripció del problema i de la solució proposada	5
3.1 CAD: <i>Computer Aided Diagnosis</i>	5
3.2 Mètodes de <i>Deep Learning</i>	6
3.2.1 Autoencoder	6
3.2.2 Xarxes Neuronals Convolucionals	7
3.2.3 Xarxes Neuronals amb dades de volumetria	8
3.3 Mètodes de <i>Machine Learning</i>	8
3.3.1 <i>Principal Component Analysis</i>	9
3.3.2 <i>Random Forests</i>	9
3.3.3 <i>Extremely Randomized Trees</i>	10
3.3.4 <i>Bagging</i>	10
3.3.5 <i>AdaBoost</i>	10
3.3.6 <i>Gradient Tree Boosting</i>	10
3.3.7 <i>K Nearest Neighbors</i>	11
3.3.8 <i>Neighborhood Components Analysis</i>	11
3.3.9 SVM	12
3.3.10 <i>Gaussian Naive Bayes</i>	12
3.3.11 <i>Gaussian Process</i>	12
3.3.12 <i>Linear Discriminant Analysis</i> i <i>Quadratic Discriminant Analysis</i>	13
3.3.13 <i>Bayesian Gaussian Mixture</i>	13
3.4 Descripció de les ferramentes software i hardware utilitzades	13
3.4.1 Tensorflow	13
3.4.2 Keras	14
3.4.3 FSL	14
3.4.4 FreeSurfer	15
3.4.5 Scikit Learn	16
3.4.6 GPU	16
4 Experimentació i resultats	19
4.1 Descripció del conjunt de dades utilitzat per a l'experimentació	19
4.1.1 Estadístiques del conjunt de dades	19
4.1.2 Tractament de les dades	19
4.1.3 Selecció de les mostres	21

4.2	Mètriques utilitzades per a l'avaluació dels resultats	22
4.3	Experiments amb mètodes de <i>Deep Learning</i>	23
4.3.1	Xarxes Neuronals Convolucionals	23
4.3.2	Xarxes Neuronals amb dades de volumetria	26
4.4	Experiments amb mètodes de <i>Machine Learning</i>	29
4.4.1	<i>Random Forests</i>	29
4.4.2	<i>Extremely Randomized Trees</i>	29
4.4.3	<i>Bagging</i>	29
4.4.4	<i>AdaBoost</i>	31
4.4.5	<i>Gradient Tree Boosting</i>	32
4.4.6	<i>K-Nearest Neighbors</i>	34
4.4.7	<i>Neighborhood Components Analysis</i>	34
4.4.8	<i>SVM</i>	36
4.4.9	<i>Gaussian Naive Bayes</i>	40
4.4.10	<i>Gaussian Process</i>	40
4.4.11	<i>Linear Discriminant Analysis</i> i <i>Quadratic Discriminant Analysis</i>	41
4.4.12	<i>Bayesian Gaussian Mixture</i>	44
5	Conclusions	45
	Bibliografia	47
<hr/>		
	Apèndix	
A	Estadístiques del conjunt de dades ADNI	51

Índex de figures

3.1	Autoencoder	7
3.2	Xarxa neuronal convolucional per a imatges 2D	8
3.3	Xarxa neuronal per a volumetria	9
3.4	Comparació KNN i NCA	11
3.5	Exemple de mixtures de gaussianes	14
3.6	Funcionament de BET	15
3.7	Funcionament de FLIRT	16
4.1	Divisió de les mostres en particions	22
4.2	<i>Accuracy</i> de la xarxa amb imatges 2D. Partició incorrecta.	24
4.3	<i>Loss</i> de la xarxa amb imatges 2D. Partició incorrecta.	24
4.4	<i>Accuracy</i> de la xarxa amb imatges 2D. Partició correcta.	25
4.5	<i>Loss</i> de la xarxa amb imatges 2D. Partició correcta.	26
4.6	ROC xarxa amb volumetria. Últim model	27
4.7	ROC xarxa amb volumetria. Model amb millor <i>accuracy</i>	28
4.8	ROC xarxa amb volumetria. Model amb menor <i>loss</i>	28
4.9	ROC <i>Random Forests</i>	30
4.10	ROC <i>Extremely Randomized Trees</i>	31
4.11	ROC <i>Bagging</i>	32
4.12	ROC <i>AdaBoost</i>	33
4.13	ROC <i>Gradient Tree Boosting</i>	34
4.14	ROC <i>K-Nearest Neighbors</i>	35
4.15	ROC <i>Neighborhood Components Analysis</i>	36
4.16	ROC SVC <i>poly</i>	37
4.17	ROC SVC <i>rbf</i>	38
4.18	ROC NuSVC <i>nu=0.01</i>	39
4.19	ROC NuSVC <i>nu=0.1</i>	40
4.20	ROC <i>Gaussian Naive Bayes</i>	41
4.21	ROC <i>Gaussian Process</i>	42
4.22	ROC <i>Linear Discriminant Analysis</i>	43
4.23	ROC <i>Quadratic Discriminant Analysis</i>	43
4.24	ROC <i>Bayesian Gaussian Mixture</i>	44
A.1	Freqüència dels valors màxims d'intensitat de les imatges ADNI	53
A.2	Freqüència dels valors màxims d'intensitat excloent valors alts de les imatges ADNI	53

Índex de taules

3.1	Comparació de les GPU	16
4.1	Resultats xarxes neuronals amb volumetria	26
4.2	Resultats xarxes neuronals amb volumetria. Model amb major <i>accuracy</i>	27
4.3	Resultats xarxes neuronals amb volumetria. Model amb menor <i>loss</i>	27
4.4	Resultats <i>Random Forest</i> en validació	29
4.5	Resultats <i>Random Forest</i> en test	29
4.6	Resultats <i>Extremely Randomized Trees</i> en validació	30
4.7	Resultats <i>Extremely Randomized Trees</i> en test	30
4.8	Resultats <i>Bagging</i> en validació	31
4.9	Resultats <i>Bagging</i> en test	31
4.10	Resultats <i>AdaBoost</i> en validació	32
4.11	Resultats <i>AdaBoost</i> en test	32
4.12	Resultats <i>Gradient Tree Boosting</i> en validació	33
4.13	Resultats <i>Gradient Tree Boosting</i> en test	33
4.14	Resultats <i>K-Nearest Neighbors</i> en validació	34
4.15	Resultats <i>K-Nearest Neighbors</i> en test	34
4.16	Resultats <i>Neighborhood Components Analysis</i> en validació	35
4.17	Resultats <i>Neighborhood Components Analysis</i> en test	35
4.18	Resultats SVC <i>poly</i> en validació	36
4.19	Resultats SVC <i>poly</i> en test	36
4.20	Resultats SVC <i>rbf</i> en validació	37
4.21	Resultats SVC <i>rbf</i> en test	37
4.22	Resultats NuSVC <i>nu=0.01</i> en validació	38
4.23	Resultats NuSVC <i>nu=0.01</i> en test	38
4.24	Resultats NuSVC <i>nu=0.1</i> en validació	39
4.25	Resultats NuSVC <i>nu=0.1</i> en test	39
4.26	Resultats <i>Gaussian Naïve Bayes</i> en validació	40
4.27	Resultats <i>Gaussian Naïve Bayes</i> en test	40
4.28	Resultats <i>Gaussian Process</i> en validació	41
4.29	Resultats <i>Gaussian Process</i> en test	41
4.30	Resultats <i>Linear Discriminant Analysis</i> en validació	42
4.31	Resultats <i>Linear Discriminant Analysis</i> en test	42
4.32	Resultats <i>Quadratic Discriminant Analysis</i> en validació	42
4.33	Resultats <i>Quadratic Discriminant Analysis</i> en test	42
4.34	Resultats <i>Bayesian Gaussian Mixture</i> en validació	44
4.35	Resultats <i>Bayesian Gaussian Mixture</i> en test	44
A.1	Recompte de sessions per edat, sexe i etiqueta	51
A.2	Recompte de subjectes per edat, sexe i etiqueta	52
A.3	Recompte de sessions per mida d'imatge 2D	52

CAPÍTOL 1

Introducció

El reconeixement d'imatges és un camp de la tecnologia que s'encarrega d'identificar persones, animals, objectes o altres elements, com patrons, gràcies a l'ús de l'aprenentatge automàtic –*machine learning*–. El terme *reconeixement d'imatges* està estretament relacionat amb el terme *visió per computador*, que és el terme utilitzat per a referir-se al procés d'entrenar els ordinadors per a poder "veure" com un humà.

Actualment, aquesta àrea de la informàtica té multitud d'aplicacions en diversos àmbits: en la indústria automobilística i els cotxes d'auto-conducció, en els videojocs i la realitat augmentada, en seguretat i reconeixement facial, o inclús en la indústria sanitària, entre d'altres. Aquesta última aplicació és la que a nosaltres ens interessa.

Quant a la introducció del reconeixement d'imatges en medicina, els principals usos que se li han donat són assistir durant el tractament o les intervencions, com, per exemple, durant les cirurgies, i ajudar a donar un diagnòstic més precís. En el nostre cas, analitzarem imatges de cervell per a crear classificadors que aprenguen a distingir cervells afectats de la malaltia Alzhèimer d'altres que no la pateixen.

1.1 Motivació i objectius

L'Alzhèimer és una malaltia neurodegenerativa caracteritzada típicament per la pèrdua de la memòria immediata i d'altres capacitats mentals a mesura que moren les cèl·lules nervioses –neurons– i s'atrofien diferents zones del cervell. Actualment, es tracta d'una malaltia incurable i terminal que apareix amb major freqüència en persones majors de 65 anys.

Alguns estudis indiquen que cada any apareixen entre 10 i 15 casos de demència per cada mil persones i, d'aquests, aproximadament la meitat són d'Alzhèimer [1]. En 2015, la Primera Conferència Ministerial de l'OMS sobre l'Acció Mundial Contra la Demència va estimar que hi havia més de 47 milions de casos de persones afectades per l'Alzhèimer al món [2].

Les proves d'imatge cerebral –tomografia axial computaritzada (TAC), ressonància magnètica nuclear (RMN), tomografia per emissió de positrons (TEP, o PET, per les seues sigles en anglés: *positron-emission tomography*) o la tomografia computaritzada per emissió de fotó únic– poden mostrar diferents signes que existeix una demència, però no especifiquen de quin tipus de demència es tracta. Per tant, el diagnòstic de la malaltia d'Alzhèimer es basa tant en la presència de certes característiques neurològiques i neuropsicològiques com en l'absència d'un diagnòstic alternatiu i es recolza en l'escàner cerebral per a detectar signes de demència. Rebre un diagnòstic precís en l'etapa primerenca de la malaltia és molt important ja que permet beneficiar-se dels tractaments

disponibles, rebre serveis de suport, participar en estudis i assajos clínics, a més de donar la possibilitat al pacient d'expressar els seus desitjos futurs i ordenar plans financers i legals [3].

L'objectiu principal d'aquest projecte és, aleshores, replicar els models de l'estat de l'art que millors resultats han aconseguit en la classificació d'imatges MRI per a, a continuació, intentar millorar els seus resultats.

1.2 Estructura de la memòria

La memòria d'aquest projecte s'estructura en cinc parts principals. Primerament, es farà un resum de l'estat de l'art actual relacionat amb la classificació d'imatges de cervell afectades per l'Alzhèimer, més específicament, d'aquells que han utilitzat el conjunt de dades ADNI. A continuació, es descriurà el problema en detall així com la solució proposada. Després, s'explicaran les ferramentes software i hardware utilitzades al llarg del projecte. Seguidament, es detallaran els experiments realitzats juntament amb el conjunt de dades utilitzat i s'acompanyaran amb un anàlisi dels resultats obtinguts. Finalment, es comentaran les conclusions a les quals s'ha arribat durant l'experimentació.

1.3 Col·laboracions

Aquest projecte ha sigut desenvolupat en col·laboració amb dos companys del grau: Álvaro López Chilet i Rafael Vicente Sánchez Romero. Una primera part del projecte ha sigut realitzada conjuntament; aquesta ha consistit principalment en l'anàlisi de les dades de les quals disposàvem, investigació de l'estat de l'art, preprocessament de les dades i uns primers experiments de classificació. Posteriorment, cadascú s'ha centrat en un tipus d'experimentació concret segons el tipus de dades utilitzat: Álvaro s'ha encarregat de les imatges en format 3D processades amb FLIRT, Rafael Vicente ha experimentat amb talls en 2D de les imatges processades amb BET i jo m'he centrat en les dades de volumetria obtingudes gràcies a FreeSurfer.

CAPÍTOL 2

Estat de l'Art

Investiguem els experiments realitzats utilitzant el conjunt de dades ADNI, format per imatges de ressonància magnètica de cervell classificades segons en quina fase de la malaltia Alzhèimer es troben els pacients. Molts articles utilitzen aquest conjunt de dades; no obstant, cadascun aplica uns mètodes de preprocessament diferents a més de proposar diferents aproximacions a la classificació de les imatges.

Tot i que les imatges del conjunt ADNI es troben classificades en cinc etiquetes diferents, la majoria d'autors se centren en aconseguir diferenciar sols entre dues d'elles: CN –*Cognitive Normal*–, que marca aquells pacients no afectats per l'Alzhèimer, i AD –*Alzheimer's Disease*–, que és l'etiqueta atorgada als pacients en fase avançada de l'Alzhèimer. D'altres opten per incloure també l'etiqueta MCI –*Mild Cognitive Impairment*–, que marca un estat intermedi entre CN i AD; alguns distingeixen en aquest grup dos subgrups: MCIc, o aquells pacients diagnosticats amb MCI que desenvolupen Alzhèimer en menys de 18 mesos, i MCInc, o aquells pacients diagnosticats amb MCI que 18 mesos després no han sigut diagnosticats amb Alzhèimer [4].

Respecte a l'aproximació escollida, trobem numerosos articles on apliquen xarxes neuronals convolucionals –CNN– per la seua adequació a problemes enfocats a trobar característiques discriminatives en imatges, així com validació creuada de 10 iteracions o classificació per votació. Als articles també s'explica la distribució d'etiquetes a les particions d'entrenament i test, i tots coincideixen a equilibrar, en certa mesura, el nombre de mostres de cada etiqueta en cadascuna de les particions de manera que no es trobe cap etiqueta de forma majoritària i puga provocar una inclinació en la classificació. No obstant, no se solen explicar les decisions que s'han pres respecte a la forma de tractar imatges diferents d'un mateix pacient. A [5] es preocupen de fer la divisió en conjunts d'entrenament i test sense que un mateix subjecte, o pacient, aparega en més d'un d'aquests conjunts; a més, proposen un mètode de classificació per votació que té en compte no només la classe més votada sinó també la que ha sigut escollida amb més probabilitat, és a dir, la predicció més certa o creïble.

Quant a preprocessament de les dades, és comú entre la majoria d'experiments normalitzar la intensitat dels píxels de les imatges i desfer-se de tot allò que no és cervell, com són el crani o els ulls, mitjançant el software BET¹ o, fins i tot, FLIRT², que, a més, estandarditza les dimensions dels cervells utilitzant-ne un model com a referència. Alguns divideixen les imatges 3D en talls 2D i apliquen transformacions sobre aquests com translació de la ROI de l'hipocamp un píxel en cadascuna de les dimensions. A [6] utilitzen les imatges 2D en projecció sagital, en lloc d'utilitzar-les en projecció axial, i utilitzen tres talls consecutius de la zona de l'hipocamp com a entrada a la xarxa neuronal en lloc

¹*Brain Extraction Tool*

²*FMRIB's Linear Image Registration Tool*

de cadascun d'ells de forma individual. Aprofitant la simetria de l'hipocamp, a més de la translació anteriorment esmentada, també li donen la volta a la imatge sobre l'eix vertical.

Trobem també investigacions centrades en les dades volumètriques que es poden obtenir a partir de les imatges de les dades ADNI; és habitual utilitzar el software FreeSurfer [7] per a extreure aquestes característiques. Algunes de les característiques que millors resultats han donat per a entrenar models que aprenquen a classificar en diverses classes són el gruix cortical, la densitat de la substància grisa, la superfície i curvatura cortical i la forma de l'hipocamp [8]. Aquestes dades també han sigut utilitzades en mètodes tradicionals de *Machine Learning* com les màquines de suport vectorial –SVM– [9] i en *Random Forests*.

2.1 Crítica a l'Estat de l'Art

Com es pot comprovar, existeix una gran quantitat d'investigacions que utilitzen el conjunt de dades ADNI per entrenar classificadors que siguin capaços de diferenciar entre pacients amb Alzhèimer i pacients sense la malaltia. La majoria mencionen les ferramentes que han utilitzat durant el preprocessament però no entren en detalls concrets d'aquest que ajudarien a replicar els experiments realitzats i a comprovar la veracitat dels resultats publicats. Als articles no sol haver-hi cap explicació sobre com han decidit la mida de les imatges utilitzades per a l'entrenament ni com han modificat les imatges originals per arribar a aquesta mida. Altra informació obviada és el repartiment de les imatges en els conjunts d'entrenament i de test que, com que hi sol haver més d'una imatge per pacient, si no es fa minuciosament, pot dur a una falsedat en els resultats.

Descripció del problema i de la solució proposada

Disposem d'un conjunt de dades obtingut gràcies a l'*Alzheimer's Disease Neuroimaging Initiative* –ADNI–, format per imatges de ressonància magnètica de cervell. Aquestes imatges es troben en format NIfTI representades en 3 dimensions i etiquetades segons l'etapa de la malaltia Alzhèimer en què es trobava el pacient al qual pertany la imatge en el moment en què aquesta es va adquirir. La classificació d'aquestes imatges ha sigut realitzada per professionals, per tant es troben correctament etiquetades, i es distingeixen cinc etiquetes diferents que són, ordenades de menor a major grau de demència: CN –*Cognitive Normal*–, EMCI –*Early Mild Cognitive Impairment*–, MCI –*Mild Cognitive Impairment*–, LMCI –*Late Mild Cognitive Impairment*– i AD –*Alzheimer's Disease*–.

Les diferències observables entre els cervells de pacients diagnosticats amb etapes diferents de l'Alzhèimer són subtils inclús entre les etapes més extremes, és a dir, entre un pacient completament sa i un altre en l'etapa més avançada de l'Alzhèimer. Aquesta falta de diferenciació es deu, principalment, al deteriorament normal del cervell degut a l'edat. Per aquest motiu, la majoria d'investigacions se centren en distingir sols entre CN i AD, o entre CN, MCI i AD. En aquest projecte també ens proposem aconseguir un classificador binari capaç de diferenciar, solament, entre AD i CN seguint les recomanacions de la nostra tutora al CIPF. Tot i això, cal destacar la importància d'un diagnòstic correcte en les primeres etapes de la malaltia. Encara que l'Alzhèimer no té cura, un tractament adequat pot resultar de gran ajuda per a mitigar els símptomes de la malaltia.

L'objectiu futur final d'aquest projecte seria, aleshores, proporcionar una ferramenta que ajude en la presa de decisions en el diagnòstic de l'Alzhèimer, especialment en els centres de la Comunitat Valenciana. Per a açò, seria necessària la col·laboració entre professionals de la salut per a reunir un conjunt de dades correctament etiquetat i suficientment extens per a poder adaptar els resultats de les nostres investigacions al tipus d'imatges amb les quals treballen els centres de la Comunitat Valenciana, que solen ser de pitjor qualitat que les del conjunt ADNI.

3.1 CAD: *Computer Aided Diagnosis*

Els diagnòstics assistits per computador, o CAD per les seues sigles en anglés, són procediments destinats a ajudar els professionals a interpretar els resultats que els pacients obtenen a les diferents proves mèdiques a les quals se sotmeten. Els sistemes que implementen CAD poden incrementar la capacitat del doctor d'arribar a un diagnòstic precís, a més de reduir el temps requerit per a tal feina.

L'àrea dels CAD és un camp interdisciplinari, on intervien tant la intel·ligència artificial i la visió per computador com el processament d'imatges radiològiques i patològiques com els raigs X, les imatges de ressonància magnètica –MRI, *magnetic resonance imaging*– o els ultrasons.

En cap cas un CAD proporciona un diagnòstic complet ni substitueix la feina del doctor o professional mèdic, sinó que, a partir de continguts multimèdia, com poden ser imatges mèdiques, marquen regions d'interès –ROI, *regions of interest*– on troben detalls que, en molts casos, no podrien haver sigut detectats pel metge per tal que aquest pugui arribar a una conclusió més clara sobre l'estat del pacient.

El procés seguit per a analitzar les dades i mostrar un resultat és sempre similar. El primer pas és el preprocessament, on es redueix el soroll, s'augmenta la qualitat i es filtren les imatges. El següent pas és la segmentació de les imatges identificant estructures –coll, cervell, teixits...–, trobant coincidències amb la base de dades anatòmica o mostrejant els valors de gris de les zones d'interès. Una vegada segmentada la imatge, s'analitzen les regions d'interès per a trobar característiques especials com compacitat, forma, mida, ubicació, proporció dels nivells de gris, etcètera. Finalment, cadascuna de les regions d'interès o imatge s'avalua individualment, ja siga utilitzant xarxes neuronals artificials, màquines de suport vectorial –SVM, *support vector machines*– o altres algorismes de classificació.

Normalment, les imatges mèdiques utilitzades en CAD es troben en format DICOM –*Digital Imaging and Communications in Medicine*–. DICOM és un estàndard de comunicació i gestió d'informació en imatge mèdica i dades relacionades amb aquesta. És un format utilitzat arreu del món que s'aplica tant en radiografia com en ecografia o MRI, entre d'altres. Un objecte DICOM conté informació que no pot ser separada de la imatge i atributs com el nom, un id, o dades de píxels.

El CAD és utilitzat en el diagnòstic de diversos càncers com poden ser el de mama, el de pulmó o el de còlon per detectar si un tumor és benigne o maligne. També és habitual en la detecció de malalties del cor per a trobar anomalies com murmurs o sons irregulars. Altra malaltia en la qual el CAD pot ser un aliat és l'Alzhèimer, per exemple, per detectar zones del cervell deteriorades.

3.2 Mètodes de *Deep Learning*

Una de les solucions proposades per a fer front al problema esmentat, consisteix en fer ús de xarxes neuronals.

3.2.1. Autoencoder

Els autoencoders són un tipus de xarxes neuronals artificials no supervisades que aprenen a comprimir i codificar dades de forma eficient i a reconstruir-les després. Aquesta reconstrucció es fa a partir de la representació codificada reduïda amb l'objectiu d'aconseguir un resultat tan paregut a les dades originals com siga possible. Els autoencoders se solen utilitzar per a detectar anomalies o per a reduir el soroll de les imatges.

Altra aplicació comú dels autoencoders consisteix a “congelar” l'estat de la xarxa en alguna de les seues capes per a utilitzar *transfer learning*, o aprenentatge per transferència. Aquesta xarxa preentrenada pot resultar d'ajuda a una segona xarxa que necessita entendre les característiques clau de les dades d'entrada per a classificar-les.

A la figura 3.1 es mostra la topologia de l'autoencoder proposat per a ser aplicat a les imatges MRI en 2D del conjunt de dades ADNI.



Figura 3.1: Autoencoder destinat al preentrenament amb les imatges ADNI en 2D per a utilitzar els seus pesos calculats com a ajuda a altres xarxes neuronals encarregades de la classificació.

3.2.2. Xarxes Neuronals Convolucionals

Per al problema consistent en classificar les imatges MRI de cervell dels pacients segons si tenen o no la malaltia d'Alzhèimer, s'ha decidit utilitzar xarxes neuronals convolucionals. Les xarxes neuronals convolucionals són un tipus de xarxa neuronal artificial comunament aplicada en l'anàlisi d'imatges visuals, o visió per computador.

Aquest tipus de xarxa assigna pesos a patrons que troba a la imatge i és capaç de capturar amb èxit les dependències espacials que hi són presents amb l'aplicació de filtres rellevants. L'objectiu de les xarxes convolucionals és reduir les imatges a una forma més fàcil de processar però sense perdre característiques que siguin fonamentals per a obtenir una bona predicció. Aquesta característica resulta important per a aconseguir una arquitectura escalable a datasets de mides considerables.

Els components més destacables de les xarxes neuronals convolucionals són les convolucions. Les convolucions són realitzades pel kernel, o filtre, sobre una part de la imatge multiplicant els elements entre sí. La mida del kernel determina sobre quants elements es realitza l'operació, i el *stride* defineix el desplaçament del kernel sobre la matriu que representa la imatge. L'aplicació d'aquests passos permet destacar parts de la imatge que ajuden a diferenciar entre les classes possibles.

A la figura 3.2 es mostra la topologia de xarxa proposada. Aquesta consta de diverses capes convolucionals seguides per quatre blocs formats per una capa de *batch normalization* i altra de *ReLU*. Finalment, la classificació serà determinada per l'eixida d'una capa amb *softmax* com a funció d'activació.

Donada la gran quantitat d'imatges i la mida d'aquestes, ha sigut necessari crear un *DataGenerator* que proporcionara a l'entrada de la xarxa les imatges dels conjunts d'entrenament i validació.

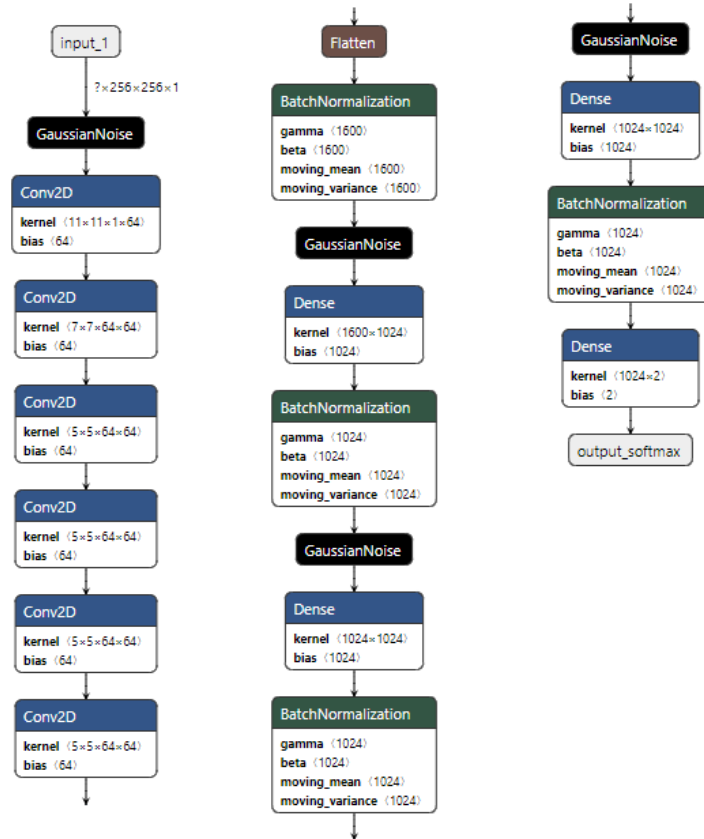


Figura 3.2: Topologia de la xarxa neuronal convolucional utilitzada per a entrenar amb imatges MRI de cervell en 2D.

3.2.3. Xarxes Neuronals amb dades de volumetria

A banda d'utilitzar imatges per a entrenar xarxes neuronals, s'ha decidit utilitzar també les dades de volumetria que calcula FreeSurfer. En aquest cas no es fan servir convolucions sinó, simplement, capes *ReLU* i *batch normalization* aplicant *dropout*. La figura 3.3 representa la xarxa proposada.

3.3 Mètodes de *Machine Learning*

L'altra solució proposada consisteix en utilitzar dades de volumetria per a crear classificadors basats en mètodes clàssics de *machine learning*.

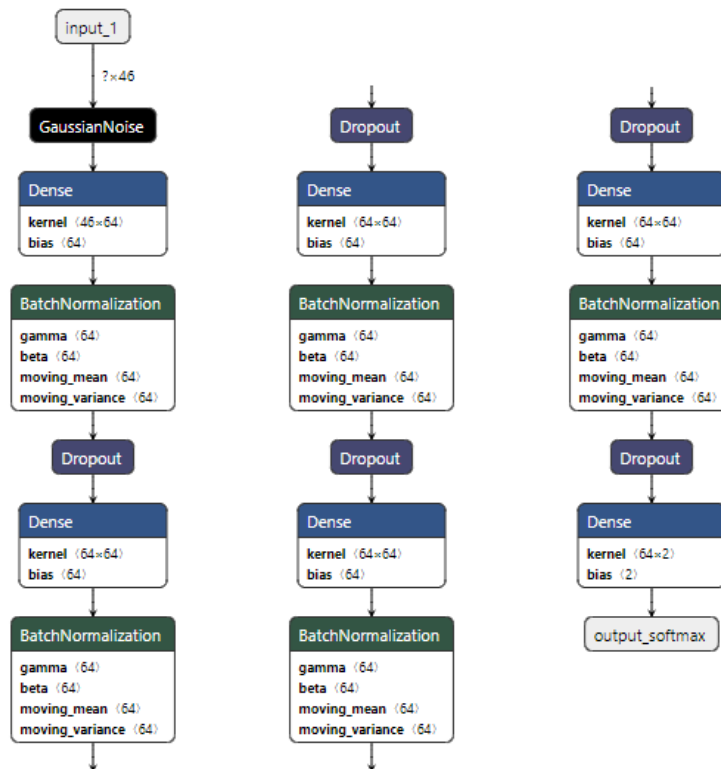


Figura 3.3: Topologia de la xarxa neuronal utilitzada per a entrenar amb dades de volumetria.

3.3.1. *Principal Component Analysis*

L'anàlisi de components principals, o PCA, és un mètode de reducció de la dimensionalitat que normalment s'utilitza per reduir la dimensionalitat dels grans conjunts de dades, transformant les mostres a un espai de menys dimensions que encara conté la informació rellevant.

Reduir el nombre de variables d'un conjunt de dades es fa a costa de la precisió, però el truc en la reducció de la dimensionalitat és intercanviar un poc de precisió per simplicitat. Un conjunt de dades de menor dimensionalitat és més fàcil d'explorar i visualitzar. A més, fa que l'anàlisi de les dades siga molt més fàcil i ràpid per als algorismes d'aprenentatge automàtic sense processar variables estranyes.

3.3.2. *Random Forests*

En els *random forests*, cada arbre del conjunt es construeix a partir d'una mostra del conjunt d'entrenament. A més, quan es divideix un node durant la construcció de l'arbre, la divisió escollida ja no és la millor divisió entre totes les funcions. En canvi, la divisió escollida és la millor divisió entre un subconjunt aleatori de les funcions. Com a resultat d'aquesta aleatorietat, la parcialitat del bosc sol augmentar lleugerament –respecte a la parcialitat d'un sol arbre no aleatori–, però, a causa de la mitjana, la seva variància també disminueix, normalment més que compensar l'augment de la parcialitat, per tant, donant un model global millor.

3.3.3. *Extremely Randomized Trees*

En els *extremely randomized trees*, l'aleatorietat va un pas més enllà en la forma en què es calculen les separacions. Igual que en els boscos aleatoris, s'utilitza un subconjunt aleatori de *candidate features*, però en comptes de buscar els llindars més discriminatius, els llindars es dibuixen aleatòriament per a cada *candidate feature* i el millor d'aquests llindars generats aleatòriament s'escull com a regla de divisió. Això normalment permet reduir la variància del model un poc més, a costa d'un augment lleugerament superior de la parcialitat.

3.3.4. *Bagging*

En els algorismes de conjunt, els mètodes de *bagging* formen una classe d'algorismes que construeixen diverses instàncies d'un estimador de caixa negra en subconjunts aleatoris del conjunt d'entrenament original i després agrupen les seves prediccions individuals per formar una predicció final. Aquests mètodes s'utilitzen com una manera de reduir la variància d'un estimador de base –per exemple, un arbre de decisió–, mitjançant la introducció d'una aleatorització en el seu procediment de construcció i després fer-ne un conjunt.

En molts casos, els mètodes de *bagging* constitueixen una manera molt senzilla de millorar respecte a un sol model, sense necessitat d'adaptar l'algorisme de base subjacent. Com que ofereixen una manera de reduir l'*overfitting*, els mètodes de *bagging* funcionen millor amb models forts i complexos –com poden ser els arbres de decisió totalment desenvolupats–, en contrast amb mètodes d'augment que normalment funcionen millor amb models febles –per exemple, arbres de decisió poc profunds–.

3.3.5. *AdaBoost*

El principi bàsic d'AdaBoost és adaptar-se a una seqüència d'aprenents febles –és a dir, models que només són un poc millors que classificar de forma aleatòria, com ara arbres de decisió menuts– en versions modificades repetidament de les dades. Les prediccions de totes elles es combinen a través d'una majoria ponderada –o suma– per a produir la predicció final.

Les modificacions de les dades en cadascuna de les anomenades iteracions de reforç consisteixen a aplicar pesos $\omega_1, \omega_2, \dots, \omega_N$ a cadascuna de les mostres d'entrenament. Inicialment, tots els pesos estan definits a $\omega_i = 1/N$, de manera que el primer pas simplement entrenarà un aprenent amb les dades originals. Per a cada iteració successiva, els pesos de la mostra es modifiquen individualment i es torna a aplicar l'algorisme d'aprenentatge a les dades de ponderació. En un pas donat, els pesos d'aquells exemples d'entrenament que el model del pas anterior va predir incorrectament augmenten, mentre que els d'aquells predits correctament es redueixen. A mesura que continuen les iteracions, els exemples difícils de predir reben una influència cada vegada més gran. D'aquesta manera, cada aprenent feble posterior es veu obligat a concentrar-se en els exemples que els anteriors en la seqüència no han sabut predir correctament.

3.3.6. *Gradient Tree Boosting*

El *gradient tree boosting* és una generalització de l'augment de funcions de pèrdua diferenciables arbitràries. El *gradient tree boosting* és un procediment eficaç i precís que es pot utilitzar tant per a problemes de regressió com per a classificació. Els seus models de reforç s'utilitzen en diferents àrees, incloent el rànquing de cerca web i l'ecologia.

Els avantatges del *gradient tree boosting* són:

- Tractament natural de dades de tipus mixt –característiques heterogènies–.
- Potència predictiva.
- Robustesa a valors externs en l'espai de sortida –mitjançant funcions de pèrdua robustes–.

El seu desavantatge és:

- L'escalabilitat, a causa de la naturalesa seqüencial de l'augment, difícilment es pot paral·lelitzar.

3.3.7. *K Nearest Neighbors*

La classificació basada en veïns és un tipus d'aprenentatge basat en instàncies o un aprenentatge no generalitzat: no intenta construir un model intern general, sinó que només emmagatzema instàncies de les dades d'entrenament. La classificació es calcula a partir d'un vot majoritari simple dels veïns més propers de cada punt: a un punt se li assigna la classe de dades que té més representants entre els veïns més propers del punt. La classificació bàsica de veïns més properes utilitza pesos uniformes: és a dir, el valor assignat a un punt de consulta es calcula a partir d'un vot de majoria simple dels veïns més propers. En algunes circumstàncies, és millor pesar els veïns, de manera que els veïns més propers contribueixin més a la forma.

3.3.8. *Neighborhood Components Analysis*

Neighborhood Component Analysis (NCA) és un algorisme d'aprenentatge automàtic per a l'aprenentatge mètric. Aprèn una transformació lineal de manera supervisada per millorar la precisió de classificació d'una regla estocàstica de veïns més propers a l'espai transformat. A la figura 3.4 podem veure la diferència que provoca utilitzar NCA en la classificació.

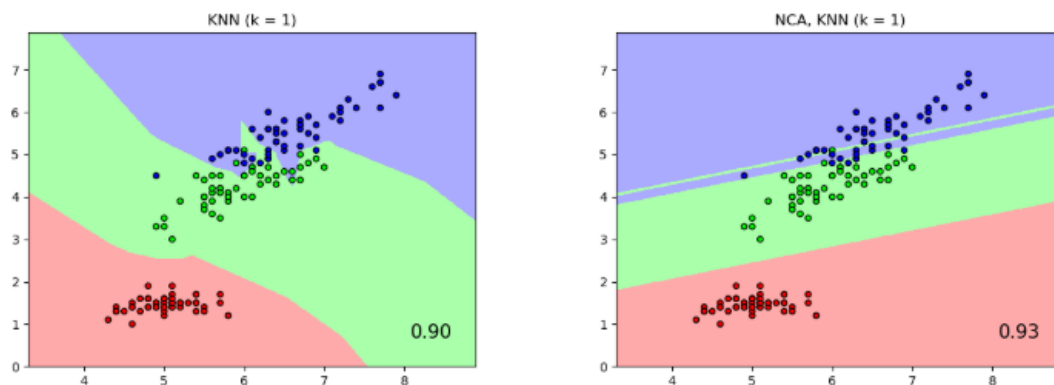


Figura 3.4: Comparació dels resultats entre una classificació basada en veïns més propers i una classificació basada en veïns més propers que a més utilitza *neighborhood components analysis* (scikit-learn.org).

3.3.9. SVM

Les màquines de suport vectorial –SVM– són un conjunt de mètodes d’aprenentatge supervisats utilitzats per a la classificació, la regressió i la detecció de valors externs.

Els avantatges de les màquines de suport vectorial són:

- Efectiu en espais d’alta dimensió.
- Continua sent eficaç en casos en què el nombre de dimensions és superior al nombre de mostres.
- Utilitza un subconjunt de punts d’entrenament a la funció de decisió –anomenats vectors de suport–, de manera que també és eficient quant a memòria.
- Versàtil: es poden especificar diferents funcions del nucli per a la funció de decisió. Es proporcionen nuclis comuns, però també és possible especificar nuclis personalitzats.

Els desavantatges de les màquines de suport vectorial inclouen:

- Si el nombre de funcions és molt més gran que el nombre de mostres, evita un ajustament excessiu en l’elecció de les funcions del kernel i el terme de regularització és crucial.
- Les SVM no proporcionen estimacions de probabilitat directament, sinó que es calculen utilitzant una costosa validació creuada de cinc iteracions.

3.3.10. *Gaussian Naive Bayes*

Els mètodes de *Naive Bayes* són un conjunt d’algorismes d’aprenentatge supervisats basats en l’aplicació del teorema de Bayes amb la suposició “ingènua” d’independència condicional entre cada parell de característiques donat el valor de la variable de classe.

Els classificadors de *Naive Bayes* han funcionat bastant bé en moltes situacions del món real, com la classificació de documents i el filtratge de correu brossa. Aquests classificadors requereixen una xicoteta quantitat de dades d’entrenament per estimar els paràmetres necessaris.

Els aprenents i classificadors de *Naive Bayes* poden ser extremadament ràpids en comparació amb mètodes més sofisticats. El desacoblament de les distribucions de característiques condicionals de classe significa que cada distribució es pot estimar independentment com una distribució unidimensional. Això, al seu torn, ajuda a pal·liar els problemes derivats de la maledicció de la dimensionalitat.

3.3.11. *Gaussian Process*

Els processos gaussians són un mètode genèric d’aprenentatge supervisat dissenyat per resoldre problemes de regressió i classificació probabilística.

Els avantatges dels processos gaussians són:

- La predicció interpola les observacions –almenys per als kernels regulars–.
- La predicció és probabilística –gaussiana– de forma que es poden calcular intervals de confiança empírica i decidir, basant-se en aquests, si cal tornar a ajustar –*online fitting, adaptive fitting*– la predicció en alguna regió d’interès.

- Versàtil: es poden especificar diferents kernels. Es proporcionen kernels comuns, però també és possible especificar kernels personalitzats.

Els desavantatges dels processos gaussians són:

- No són escassos, és a dir, utilitzen tota la informació de mostres/funcions per realitzar la predicció.
- Perden eficiència en espais d'alta dimensió, és a dir, quan el nombre de característiques és superior a algunes desenes.

3.3.12. *Linear Discriminant Analysis* i *Quadratic Discriminant Analysis*

L'anàlisi discriminant lineal i l'anàlisi discriminant quadràtic són dos classificadors clàssics, amb una superfície de decisió lineal i quadràtica, respectivament. Aquests classificadors són atractius perquè tenen solucions de forma tancada que es poden calcular fàcilment, són intrínsecament multiclasse, han demostrat funcionar bé a la pràctica i no tenen hiperparàmetres per sintonitzar.

3.3.13. *Bayesian Gaussian Mixture*

Un *gaussian mixture model* és un model probabilístic que assumeix que tots els punts de dades es generen a partir d'una barreja d'un nombre finit de distribucions gaussianes amb paràmetres desconeguts. Es pot considerar que els *mixture models* generalitzen agrupaments de *k-means* per incorporar informació sobre l'estructura de covariància de les dades, així com els centres de les gaussianes latents. Podem veure un exemple de mixtura de gaussianes a la figura 3.5.

En una configuració bayesiana, els pesos i els paràmetres de la mixtura seran variables aleatòries i les distribucions anteriors se situaran sobre les variables. En aquest cas, els pesos normalment es veuen com un vector aleatori K -dimensional dibuixat a partir d'una distribució de Dirichlet¹, i els paràmetres es distribuïran segons els seus respectius priors conjugats.

3.4 Descripció de les ferramentes software i hardware utilitzades

3.4.1. Tensorflow

TensorFlow és un sistema d'aprenentatge automàtic que opera a gran escala i en entorns heterogenis. TensorFlow utilitza gràfics de flux de dades per representar la computació, l'estat compartit i les operacions que muten aquest estat. Mapifica els nodes d'un gràfic de flux de dades a través de moltes màquines d'un mateix clúster, i dins d'una màquina a través de múltiples dispositius informàtics, incloent CPU, GPU i TPU [10, 11].

TensorFlow és compatible amb una gran varietat d'aplicacions, enfocat en l'entrenament i la inferència en xarxes neuronals profundes. TensorFlow va ser llançat com a projecte de codi obert per Google, diversos serveis del qual l'empren, i s'utilitza àmpliament per a la recerca d'aprenentatge automàtic.

¹Distribució de probabilitat multivariant contínua parametritzada per un vector α de símbols reals positius.

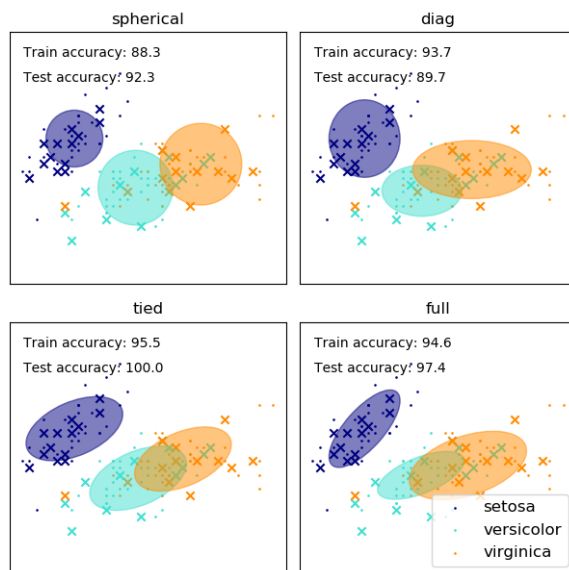


Figura 3.5: Exemple de mixtures de gaussianes (scikit-learn.org).

3.4.2. Keras

Per construir i entrenar models amb TensorFlow, la millor opció, i la més fàcil, és utilitzar l'API d'alt nivell Keras. Aquesta API no només és simple i consistent, sinó que també redueix el nombre d'accions requerides a l'usuari per implementar els casos d'ús més comuns. Gràcies a aquestes característiques, Keras se centra en permetre una ràpida experimentació. Keras està escrit en Python i pot executar per sobre de TensorFlow, CNTK o Theano. A més, admet tant xarxes convolucionals com recurrents i també combinacions d'aquestes [12].

Keras és el *framework* d'aprenentatge profund més utilitzat tant a la indústria com a la comunitat investigadora sols per darrere del propi TensorFlow.

3.4.3. FSL

FSL –*FMRIB² Software Library*– és una biblioteca de ferramentes d'anàlisi d'imatges cerebrals FMRI, MRI i DTI. Proporciona ferramentes que permeten realitzar multitud de tasques com poden ser segmentació del cervell segons el tipus de teixit –substància grisa, substància blanca...– o preprocessament de les imatges [13, 14].

Les funcionalitats d'FSL utilitzades en aquest projecte han sigut BET i FLIRT, destinades a segmentació i a registre d'imatge.

BET

BET –*Brain Extraction Tool*– és una ferramenta disponible de forma gratuïta com a part del software FSL que permet extraure la part corresponent al cervell a partir d'imatges T1 i T2, tot i que també funciona sols amb imatges T1 encara que amb menys precisió. BET es basa en la intensitat dels vòxels de la imatge per trobar el crani i el cuir cabellut i poder crear una màscara que permeta eliminar-los de la imatge. A més a més, també ofereix les opcions d'eliminar els ulls i el coll amb cert grau de fidelitat, encara que amb un cost computacional i temporal major [15, 16].

²The Oxford Centre for Functional MRI of the Brain

A fi de llevar informació, a priori, irrellevant per a la xarxa, s'ha utilitzat el software BET per a excloure el crani, els ulls i el coll de cadascuna de les imatges del conjunt de dades ADNI. Aquest tractament ha permès, també, reduir la mida de les imatges ja que molts dels talls de cadascuna de les imatges 3D pertanyien al coll.

A la figura 3.6 es poden observar les modificacions que BET és capaç de realitzar sobre una imatge MRI de cervell.

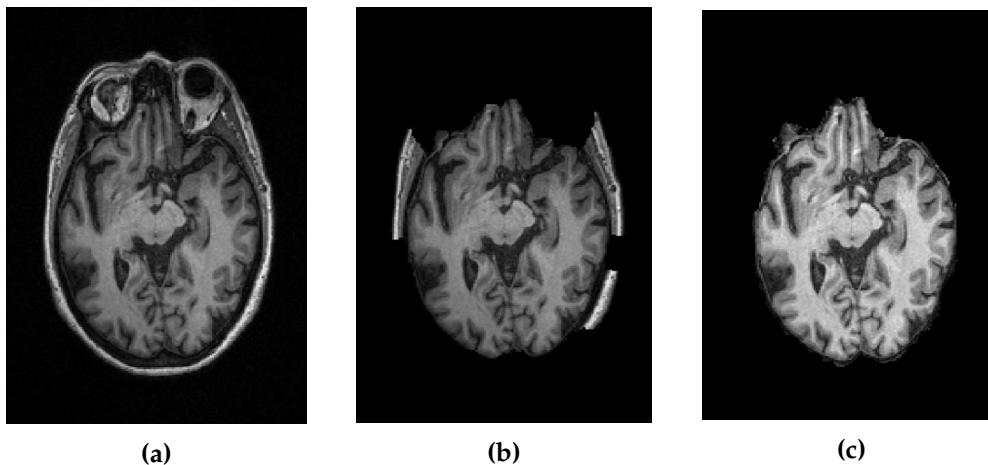


Figura 3.6: Funcionament de BET. A aquestes imatges podem veure el mateix tall 2D d'una imatge 3D. A la figura (a) es mostra la imatge original; a (b) es mostra el resultat de processar la imatge original amb BET amb el paràmetre *fractional* amb valor 0.3; a (c) podem veure el resultat de processar la imatge original amb BET utilitzant els paràmetres proposats a [17], que són *fractional*=0.1, *remove_eyes*=True i *bias*=True.

FLIRT

FLIRT –*FMRIB's Linear Image Registration Tool*– és una ferramenta robusta i precisa totalment automatitzada per al registre lineal –afí– i intermodal d'imatges del cervell. A partir d'una imatge de referència o model que represente un cervell estàndard, FLIRT s'encarrega d'ajustar la resta d'imatges del conjunt perquè totes elles mostren els cervells amb la mateixa forma i dimensions 3.7. Amb la transformació que proporciona FLIRT, s'obté un conjunt de dades més homogeni, on les característiques de cada cervell es poden veure en proporció a l'estàndard utilitzat [18, 19, 20].

3.4.4. FreeSurfer

FreeSurfer és un conjunt de ferramentes software destinades a l'estudi de l'anatomia cortical i subcortical. Aquestes s'encarreguen de construir models de la frontera entre la matèria blanca i la matèria grisa cortical, així com la superfície pial. En el moment en què es coneixen aquestes superfícies, es calculen mesures com el gruix cortical o la curvatura a cadascun dels punts de l'escorça cerebral.

La parcel·lació del cervell es realitza segons un atlas elegit, el qual ha sigut construït a partir d'un conjunt d'entrenament format per un grup de subjectes els cervells dels quals han sigut etiquetats a mà, tant les superfícies com els volums. Aquestes etiquetes són mapificades en un espai comú –a un espai MNI305 per a volums i a un espai esfèric per a superfícies–. Una vegada mapificades les etiquetes, es calculen les probabilitats per a cadascun dels vòxels de pertànyer a una etiqueta o altra. Aquest procediment s'ha demostrat estadísticament indistingible dels indicadors manuals i relativament insensible als canvis en els paràmetres d'adquisició [21, 22].

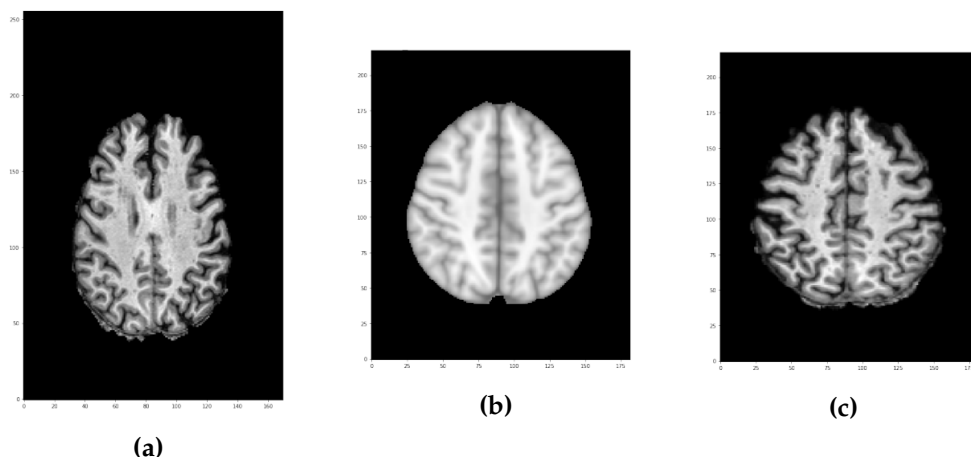


Figura 3.7: Funcionament de FLIRT. La imatge (a) correspon a un tall d’una de les imatges del conjunt original processada amb BET, la imatge (b) és el model de cervell utilitzat com a referència per a transformar la imatge original i la imatge de (c) és el resultat de la transformació. Les tres imatges ací mostrades corresponen al tall que es troba a la mateixa posició relativa en les tres imatges 3D de les quals provenen, però, la imatge (a) mostra el tall número 166 d’una imatge 3D formada per 256 talls mentre que les altres dues corresponen al tall 118 d’una imatge 3D formada per 182. A aquestes imatges podem veure, per tant, com canvien les tres dimensions de les imatges originals en utilitzar FLIRT.

3.4.5. Scikit Learn

Scikit-learn és una biblioteca d’aprenentatge automàtic de software lliure. Aquesta biblioteca està destinada al llenguatge de programació Python i, a més, està dissenyada per a poder interoperar amb les biblioteques NumPy i SciPy. Scikit-learn conté diversos algorismes de classificació, regressió, *clustering*, reducció de la dimensionalitat, selecció de models i preprocessat [23].

3.4.6. GPU

Les GPU utilitzades han estat disponibles per a aquest projecte gràcies a la col·laboració tant amb el Centre d’Excel·lència i Innovació Tecnològica de Bioimatge –CEIB–, la unitat mixta d’imatge biomèdica de la Fundació per al Foment de la Investigació Sanitària i Biomèdica –FISABIO– i el Centre d’Investigació Príncep Felip –CIPF–, com amb el centre d’investigació Pattern Recognition and Human Language Technology –PRHLT–.

S’ha disposat de dues NVIDIA GeForce RTX 2080, ubicades a la mateixa màquina, tres NVIDIA GeForce GTX 1080, ubicades cadascuna en una màquina diferent, i d’una NVIDIA Quadro P5000.

Taula 3.1: Comparació de les GPU

	RTX 2080	GTX 1080	Quadro P5000
CUDA cores	2944	2560	2560
Memòria estàndard	8 GB	8 GB	16 GB

La característica més rellevant de les GPU ha sigut la memòria disponible, que permetia treballar amb *batches* i models de major o menor mida segons si s’utilitzava la targeta amb 16 GB o alguna de les de 8 GB. A més, cal destacar la importància de poder utilitzar

GPU en lloc de CPU, especialment per a treballar amb imatges, gràcies a la seua velocitat de processament.

Experimentació i resultats

4.1 Descripció del conjunt de dades utilitzat per a l'experimentació

Per a dur a terme els experiments s'han utilitzat imatges de ressonància magnètica –MRI– de l'Alzheimer's Disease Neuroimaging Initiative –ADNI– [24]. ADNI va començar en 2004 amb l'objectiu de comprovar si les imatges MRI, les imatges PET –*positron emission tomography*– i altres biomarcadors podien ser combinats per a mesurar el progrés del deteriorament cognitiu lleu –*mild cognitive impairment*, MCI– i la malaltia Alzhèimer –*Alzheimer's disease*, AD–.

Amb la determinació de marcadors específics de la progressió de l'Alzhèimer es pre-tén ajudar els investigadors i personal clínic a desenvolupar nous tractaments i controlar la seua eficàcia, a més de disminuir el temps i el cost dels assajos clínics[25].

4.1.1. Estadístiques del conjunt de dades

El conjunt de dades que hem utilitzat està format per imatges MRI tant d'homes com de dones que tenen entre 50 i 90 anys. En total hem disposat de 2576 sessions corresponents a 1476 subjectes diferents, entenent com pertanyents a un subjecte diferent les sessions d'un mateix pacient etiquetades en categories diferents. Per tant, per cada subjecte existeixen una mitjana d'1.75 sessions.

Aproximadament, el 46% de les sessions i subjectes corresponen a dones i el 54% a homes, disposant així d'un conjunt bastant equilibrat quant a sexe dels pacients dels quals provenen les mostres. Quant a rang d'edat, el 80% de les sessions i dels subjectes corresponen a pacients d'entre 66 i 85 anys. Aquestes dades es troben de forma detallada a les taules A.1 i A.2 de l'apèndix A.

Les imatges de cadascuna de les sessions es troben en format 3D però, a causa de l'ús de maquinària heterogènia per a l'adquisició d'aquestes imatges, les mides de les imatges són molt diverses. Tot i això, cap de les imatges supera els 256 píxels en cap de les seues dimensions.

4.1.2. Tractament de les dades

Com a pas previ a la realització de cadascun dels experiments, s'han provat diverses tècniques de preprocessament de les imatges des de normalització de les intensitats fins a extracció de dades volumètriques. La complexitat d'aquests preprocessaments ha anat

augmentant segons es realitzaven més proves per poder trobar aquells que millor resultat proporcionaven.

La modificació de les imatges més simple ha sigut dividir cada imatge 3D en altres en 2D. Açò s'ha aconseguit dividint cadascuna de les imatges 3D en talls 2D de manera que les imatges resultants foren imatges en projecció axial del cervell. A partir d'aquestes imatges resultants s'ha fet un estudi de les seues mides per a decidir les dimensions a les quals transformar totes les imatges 2D per a que tinguen la mateixa mida i poder servir com a entrada a xarxes neuronals convolucionals. A la taula A.3 de l'apèndix A es pot veure el recompte de sessions per mida d'imatge 2D.

Quant a homogeneïtzació de les mostres, s'han aplicat dos transformacions: obtenir, a partir de les imatges originals en 2D, un conjunt de mostres totes amb les mateixes dimensions i normalitzar les intensitats. A la gràfica A.1 es pot veure un histograma amb la freqüència del valor de píxel màxim de cada imatge 2D, i a la gràfica A.2 es pot veure l'histograma amb les intensitats més altes excloses; és a dir, per a cada mostra 2D s'ha registrat el valor del píxel amb major intensitat.

La mida escollida per a les mostres ha sigut 176x240. En aquelles imatges que no tenien aquestes dimensions s'han aplicat les següents transformacions:

- si la primera dimensió és menor a 176 i/o la segona és menor a 240, s'ha aplicat *padding* zero fins arribar a aquestes dimensions, deixant la imatge original centrada en la imatge final
- si la primera dimensió és major a 176 o la segona és major a 240, s'obtenen, a partir de la imatge, dos imatges diferents de 176x240, cadascuna corresponent a un extrem de la imatge en la dimensió que excedeix el valor escollit
- si la primera dimensió és major a 176 i la segona és major a 240, s'obtenen, a partir de la imatge, cinc imatges diferents de 176x240, quatre d'elles corresponents a cadascun dels cantons de la imatge i una altra corresponent al centre de la imatge

Per a normalitzar les intensitats s'ha optat per obviar aquells valors molt més alts que la resta de màxims, ja que, probablement, corresponen a errors en l'adquisició de les imatges MRI. Així doncs, s'ha escollit com a màxim global el valor 2000 i s'han normalitzat les imatges respecte d'aquest màxim de manera que els valors d'intensitat dels píxels queden en el rang [0,1]. També s'ha provat a normalitzar cadascuna de les imatges a partir de la següent fórmula:

$$\text{nou valor del píxel} = \frac{\text{valor actual del píxel} - \text{mitjana dels valors de la imatge}}{\text{desviació estàndard dels valors de la imatge}}$$

Els dos mètodes de normalització de les intensitats de les imatges donen resultats similars en la classificació.

També s'ha provat a processar les imatges 3D amb el *software* BET 3.4.3 per a eliminar tot allò de les imatges que no correspon al cervell. D'aquesta manera s'han eliminat els ulls, coll i crani, estructures que no afecten al diagnòstic. En eliminar el coll, part de cada imatge 3D ha pogut ser eliminada ja que no contenia informació i així s'ha reduït la mida de les imatges. A partir d'aquestes imatges 3D processades en BET també s'han pogut obtenir imatges 2D seguint la mateixa metodologia explicada anteriorment.

Altres *software* utilitzat sobre les imatges originals en 3D ha sigut FLIRT 3.4.3. FLIRT adapta les imatges del cervell seguint un model de referència de manera que totes les imatges tinguen les mateixes dimensions i que la representació del cervell tinga la mateixa forma. Amb açò s'aconsegueix uniformitat i no cal redimensionar les mostres.

Per a extraure dades volumètriques s'ha emprat el *software* FreeSurfer 3.4.4. FreeSurfer s'encarrega de fer una anàlisi de les zones del cervell per dividir-lo en seccions i calcular dades com la curvatura o el volum a partir de diversos atlas disponibles. En processar les imatges 3D amb FreeSurfer s'han generat més de 70 arxius diferents amb medicions tant de l'hemisferi dret com de l'esquerre.

4.1.3. Selecció de les mostres

Per poder avaluar els resultats dels experiments de forma correcta i fiable, aquesta avaluació s'ha de realitzar sobre mostres no vistes pel sistema prèviament. Per aquest motiu, el conjunt inicial de mostres s'ha de dividir en tres conjunts diferents: entrenament, validació i test.

El conjunt d'entrenament és utilitzat per a entrenar el model –ja siga un classificador de Bayes o una xarxa neuronal, per exemple– utilitzant un mètode d'aprenentatge supervisat i així adaptar els paràmetres del model. Amb el model obtingut, es prediuen les respostes per al conjunt de validació i s'aconsegueix una avaluació imparcial d'aquest model que pot donar informació sobre possibles *overfittings*. Finalment, el conjunt de test proporciona una valoració imparcial del model final adaptat a les dades d'entrenament. Per a la nostra investigació s'ha optat per dividir les mostres de forma que aproximadament un 60% d'aquestes formen part del conjunt d'entrenament, un 20% del de validació i el 20% restant del de test.

Durant els primers experiments realitzats, la divisió en aquests tres conjunts s'ha fet sobre els talls 2D. És a dir, l'únic criteri a seguir ha consistit en no repetir cap imatge 2D en més d'un conjunt. Els resultats d'aquests experiments han resultat quasi perfectes i s'ha arribat a la conclusió de que els talls consecutius d'una mateixa imatge presenten massa similituds entre ells. Per aquest motiu, en cas d'aparèixer al conjunt de test un tall corresponent a la mateixa imatge 3D que un altre tall que es troba en el conjunt d'entrenament, la xarxa el classificava amb tota seguretat a la classe correcta. A més a més, en aquests primers experiments, s'han utilitzat les cinc etiquetes, cosa que feia més sorprenents els resultats tan alts.

A causa d'aquestes proves, s'ha decidit dividir el conjunt inicial en els tres conjunts mencionats utilitzant pacients en lloc de mostres. És a dir, evitant que les mostres pertanyents a un mateix pacient apareguen en més d'un conjunt. Amb açò s'evita que el model reconega en una mostra del conjunt de test l'estructura pròpia del pacient a la qual pertany en lloc de les característiques que defineixen l'etapa de l'Alzhèimer en què es troba.

Després d'haver canviat l'estructura dels conjunts d'entrenament, validació i test, els resultats obtinguts han empitjorat prou, per la qual cosa s'ha decidit utilitzar sols les etiquetes CN i AD, tal i com és habitual en les investigacions de l'estat de l'art. A l'estat de l'art els resultats arriben al 90% d'*accuracy* –i fins i tot el superen–; no obstant, als articles no s'especifica que no es repetisquen pacients entre els conjunts. A la figura 4.1 es mostra un diagrama que compara la forma de dividir les mostres que utilitzen a l'estat de l'art amb la forma de dividir les mostres que s'ha decidit utilitzar en aquest projecte per a que els resultats siguin més fidels a l'actuació dels models en un entorn real. Aquesta partició proposada s'ha utilitzat tant a les proves amb imatges com a les proves amb dades de volumetria.

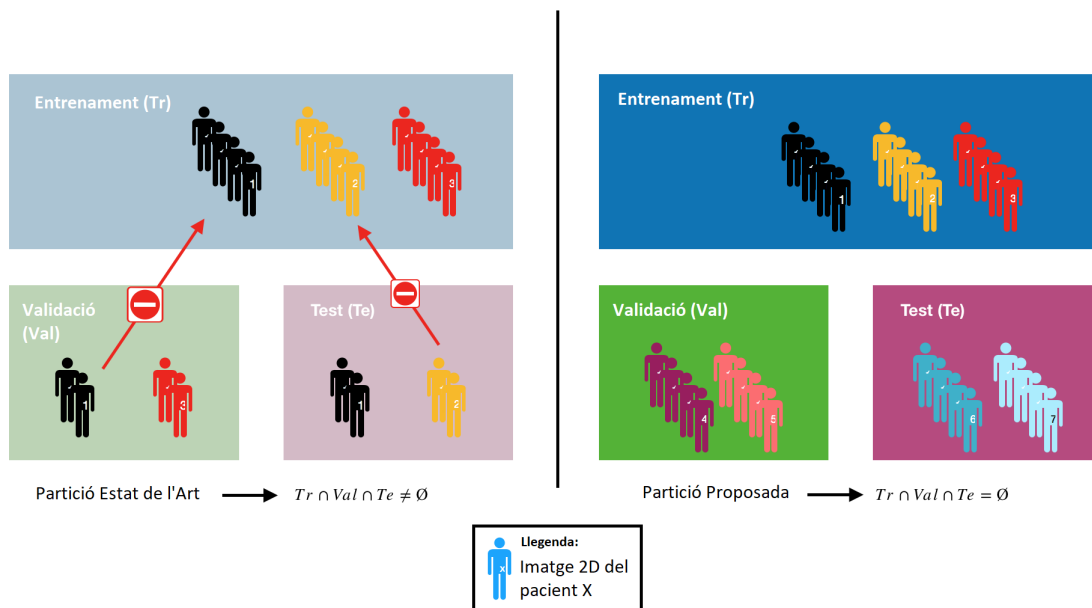


Figura 4.1: Divisió de les mostres en particions d'entrenament, validació i test de forma que un mateix pacient no aparega en més d'una d'aquestes particions.

4.2 Mètriques utilitzades per a l'avaluació dels resultats

Per a avaluar els resultats obtinguts als experiments realitzats, s'ha decidit comparar no només *accuracy* sinó també *precision*, *recall* i *f1-score*. D'aquesta forma podrem tindre una visió més exacta dels resultats. Per a definir cadascuna d'aquestes mesures parlarem de *true positives* –TP, mostres classificades en una classe que sí pertanyen a eixa classe–, *false positives* –FP, mostres classificades en una classe però que no pertanyen a eixa classe–, *true negatives* –TN, mostres no classificades en una classe que no pertanyen a eixa classe– i *false negatives* –FN, mostres no classificades en una classe però que sí pertanyen a eixa classe–. A més a més, també es calcularà l'àrea sota la corba ROC –*receiver operating characteristic*, la qual mostra la relació entre l'índex de TP i l'índex de FP.

Quan es mesura *accuracy* es fa un recompte d'aquelles mostres que han sigut correctament classificades, és a dir, *accuracy* és la proporció de resultats correctes sobre el total de casos examinats:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Aquest valor serà el mateix siga quina siga la classe sobre la qual es calcula.

Mentre que *accuracy* és una mesura global per a tota la classificació, *precision*, *recall* i *f1-score* depenen de la classe, és a dir, cadascuna de les classes tindrà un valor diferent, i la combinació d'ells ens dirà com de bé s'ha realitzat la distinció entre les classes.

Amb *precision* es calcula quantes mostres de les que han sigut classificades en un grup realment pertanyen a aquest i la seua equació és la següent:

$$Precision = \frac{TP}{TP + FP}$$

En el nostre problema es traduiria com la proporció de gent diagnosticada amb Alzhèimer que realment té Alzhèimer. Per tant, és interessant trobar un valor de *precision* per a la classe AD com més alt possible per a assegurar-nos de que tothom amb Alzhèimer és tractat com a tal.

Per altra banda tenim *recall*, que calcula quantes de les mostres rellevants per a una classe han sigut seleccionades, és a dir, quin total de les mostres pertanyents a una classe s'han arribat a classificar en eixa classe, i la seua equació és la següent:

$$\text{Recall} = \frac{TP}{TP + FN}$$

En el nostre cas, amb *recall* mesurariem, si parlem de la classe AD, la proporció de casos d'Alzhèimer que hem arribat a detectar sobre el total de casos d'Alzhèimer amb què ens hem trobat.

El *f1-score* proporciona una mitjana harmònica de *precision* i *recall* que ens permet avaluar les dues mètriques anteriors en conjunt. La seua equació és:

$$Fscore = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

El valor desitjat en *f1-score* és 1.

La corba ROC és la representació de la ràtio de vertaders positius –TP– front a la ràtio de falsos positius –FP– segons varia el llindar de discriminació o valor a partir del qual considerem un cas com positiu. L'anàlisi de la corba ROC ajuda a discriminar models possiblement òptims d'aquells subòptims. A més, la corba ROC és **independent de la distribució de les classes** en la població. Si parlem de diagnòstic, direm que la corba és independent de la prevalença d'una malaltia en la població.

4.3 Experiments amb mètodes de *Deep Learning*

Donada la dificultat del problema, s'ha decidit no utilitzar les mostres de totes les etiquetes sinó solament aquelles que pertanyen al grup CN o al grup AD, tal i com és habitual a les investigacions de l'estat de l'art. Per tant, el problema es redueix a una classificació binària.

4.3.1. Xarxes Neuronals Convolucionals

Com s'ha explicat en capítols anteriors, els primers experiments han sigut realitzats sobre les imatges MRI en 2D. La xarxa utilitzada és la descrita en la secció 3.2.2 i representada en la figura 3.2. En les primeres proves, els tres conjunts d'entrenament, validació i test no estaven ben definits, ja que imatges pertanyents a un mateix subjecte podien aparèixer en més d'un d'aquest tres conjunts.

A la figura 4.2 es mostra l'evolució de l'*accuracy*, tant del conjunt d'entrenament com de validació, al llarg de les *epochs* realitzades durant l'entrenament i comparant-la amb el resultat final del conjunt de test. S'observa una millora constant en la classificació tant de les dades d'entrenament com de les de validació, tot i que per a aquest segon grup l'*accuracy* descendeix de forma abrupta en algunes *epochs*. Resulta sorprenent com, en tot moment, l'*accuracy* del conjunt de validació es troba molt per damunt de la del conjunt d'entrenament i com la del conjunt de test és del 95%.

Els valors de *loss* al llarg de les *epochs* es mostren a la figura 4.3. Aquests valors són coherents amb els resultats d'*accuracy* representats a la figura anterior. El valor de *loss* en el conjunt d'entrenament disminueix progressivament, mentre que el de validació ho fa de forma més ràpida amb augments abruptes en *epochs* concretes, coincidint amb els pics d'*accuracy*. El valor de *loss* del conjunt de test és mínim, no superant el 0.15.

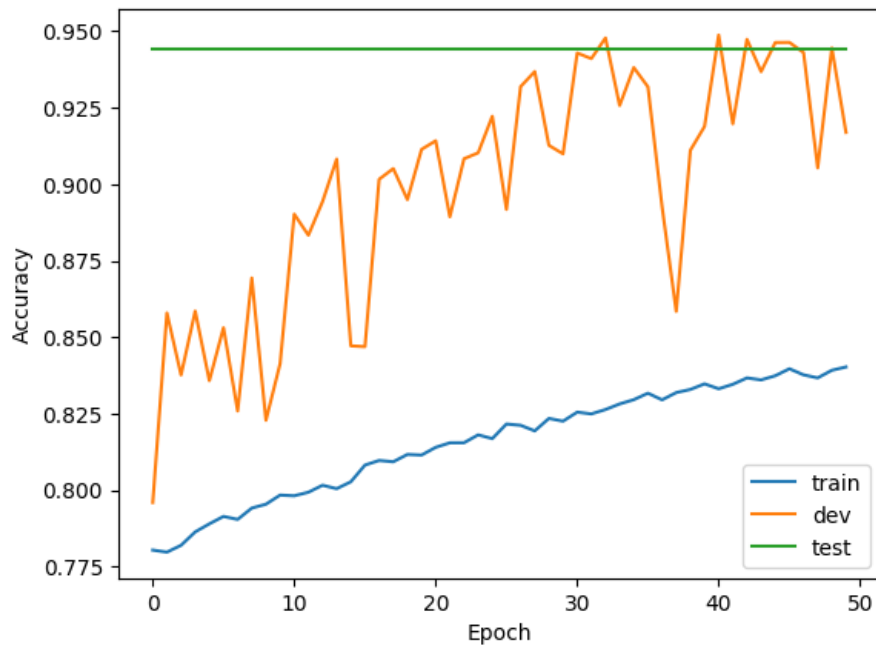


Figura 4.2: *Accuracy* obtinguda amb la xarxa neuronal convolucional en classificació d'imatges en 2D utilitzant la partició de les dades incorrecta. Com es pot veure, l'*accuracy* en el conjunt de test arriba a quasi el 95%, mentre que en el conjunt d'entrenament el màxim no arriba al 85%.

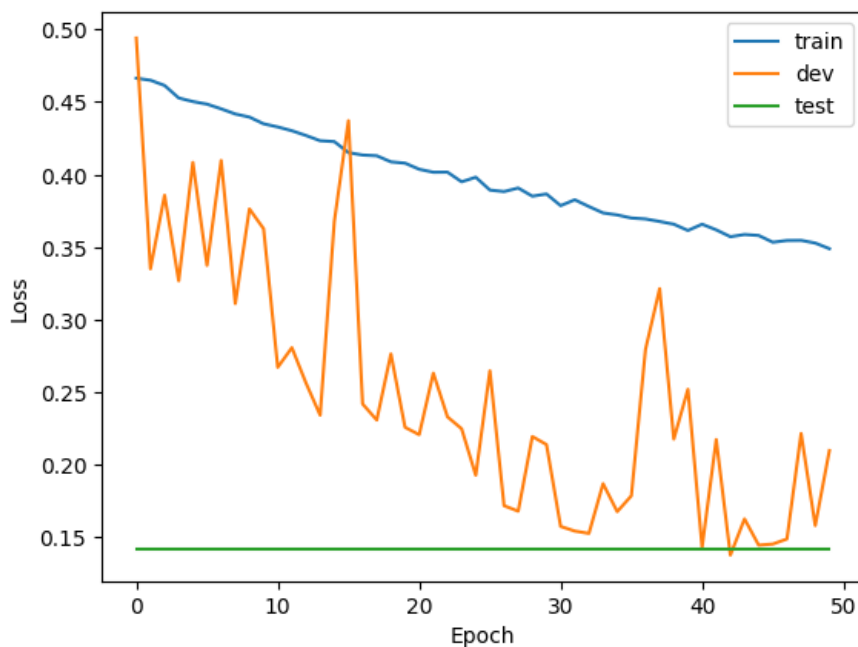


Figura 4.3: *Loss* obtingut amb la xarxa neuronal convolucional en classificació d'imatges en 2D utilitzant la partició de les dades incorrecta. Com es pot veure, el *loss* en el conjunt de test queda per baix de 0.15, mentre que en el conjunt d'entrenament disminueix fins quedar un poc per damunt de 0.35.

A partir d'aquests resultats sorprenents en el conjunt de test, s'arriba a la conclusió de que existeix un problema en el repartiment de les dades destinades a entrenar i aquelles destinades a comprovar l'eficàcia del model proposat. Per això, partint de la suposició de que totes les imatges d'un mateix pacient, les quals en la majoria dels casos tenen assignada la mateixa etiqueta, són massa similars entre sí, es crea una nova divisió de les mostres en tres conjunts. Cada pacient és assignat a un dels tres conjunts de mostres i totes les imatges pertanyents a ell formaran part del mateix grup, sempre respectant, de forma aproximada, la proporció de mostres que es trobaran en cada conjunt: 60% en entrenament, 20% en validació i 20% en test. D'aquesta manera, quan el model analitzi una imatge del conjunt de test, no podrà relacionar-la amb cap altra dels conjunts d'entrenament o de validació.

El mateix model de xarxa representat a la figura 3.2 és tornat a entrenar amb la nova partició de dades. Els resultats d'*accuracy* d'aquest experiment es poden veure a la figura 4.4. En aquest cas, la xarxa memoritza ràpidament les mostres d'entrenament, ja que, com es pot observar, a partir de l'*epoch* 20 el valor de l'*accuracy* ja no només no baixa del 95%, sinó que s'aproxima de forma asimptòtica al 100%. En canvi, l'*accuracy* per a validació i test es queda al voltant del 67%.

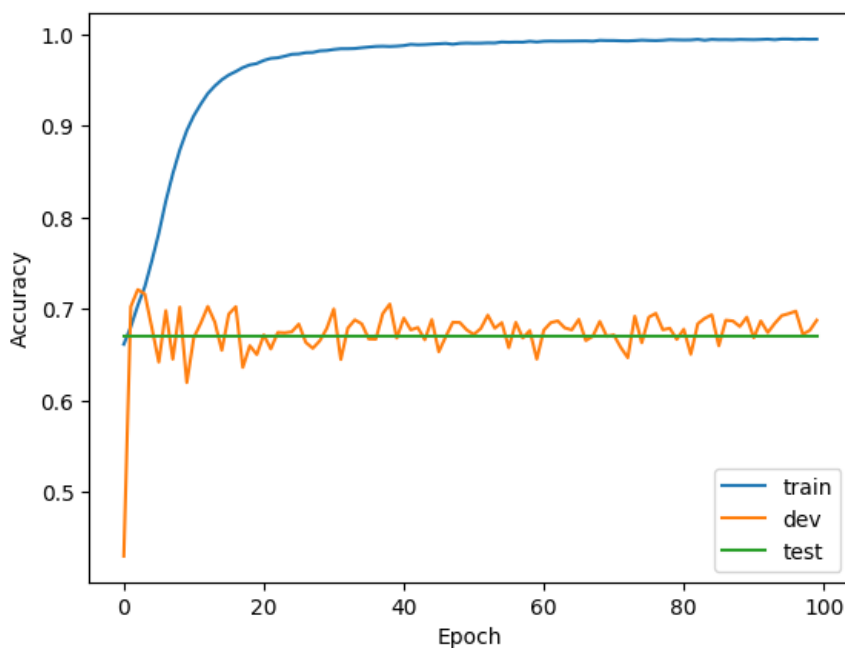


Figura 4.4: *Accuracy* obtinguda amb la xarxa neuronal convolucional en classificació d'imatges en 2D utilitzant la partició de les dades correcta. Com es pot veure, l'*accuracy* en el conjunt d'entrenament arriba a quasi el 100%, mentre que en els conjunts de validació i test l'*accuracy* es queda sobre el 67%.

A la figura 4.5 es mostra l'evolució dels valors de *loss* en aquest experiment. El valor en el conjunt d'entrenament s'aproxima al 0 absolut a mesura que es realitzen més *epochs*. El resultat en test es troba lleugerament per damunt del 0.5 però, encara que no és un mal resultat, si es té en compte el continu augment del valor de *loss* en validació, l'experiment no pot ser avaluat positivament. Aquesta gràfica i la que mostra l'*accuracy* evidencien el sobreentrenament que s'ha produït a la xarxa, per la qual cosa serà necessari adoptar mesures per a evitar-lo, com experimentar amb el *learning rate* o amb *dropout* o aplicar mètodes de preprocessament com els que proporcionen les ferramentes explicades al capítol 3.4.

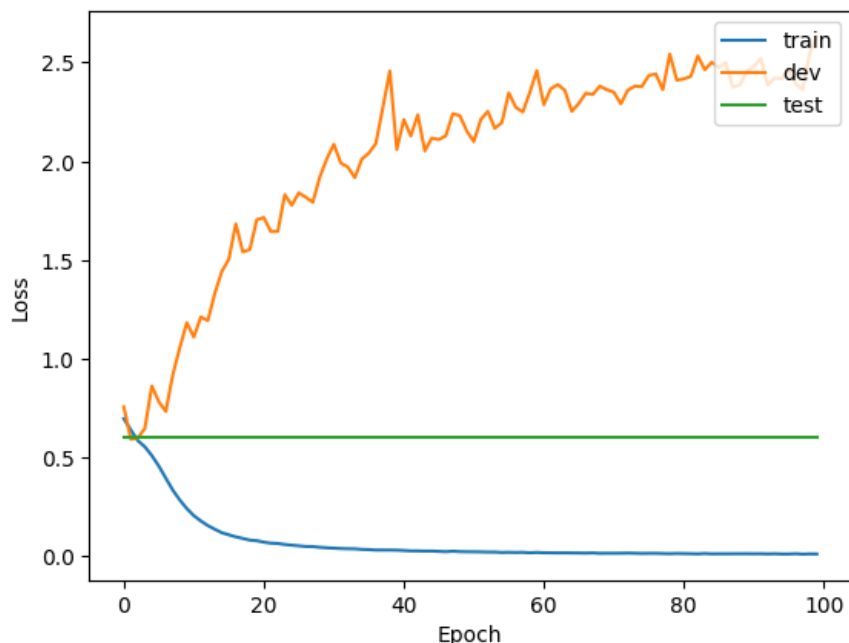


Figura 4.5: *Loss* obtingut amb la xarxa neuronal convolucional en classificació d'imatges en 2D utilitzant la partició de les dades correcta. Com es pot veure, el *loss* en el conjunt d'entrenament disminueix fins arribar a un valor molt proper a 0, mentre que en el conjunt de validació no deixa d'augmentar i sobrepassa de 2.5 al cap de 100 *epochs*.

4.3.2. Xarxes Neuronals amb dades de volumetria

Per a obtenir dades de volumetria s'ha fet ús de la ferramenta FreeSurfer, explicada en la secció 3.4.4. Donat el gran volum de medicions obtingudes gràcies a FreeSurfer, s'ha decidit aplicar PCA sobre les mostres per a reduir les dimensions d'aquestes. S'ha decidit escollir el mínim nombre de components principals de forma que es mantinga el 95% de la variància. El número total de components principals utilitzats és 46.

El model de xarxa utilitzat ha sigut el model representat a la figura 3.3. En aquest cas, s'han entrenat 10 xarxes. Per a cadascun dels 10 entrenaments realitzats s'han guardat tres models diferents: l'últim model generat per la xarxa, el model que major *accuracy* ha aconseguit en el conjunt de validació i el model que menor *loss* ha aconseguit en validació. Finalment, s'han combinat les prediccions de test dels 10 últims models generats per cada xarxa, les dels 10 models amb millor *accuracy* i les dels 10 models amb menor *loss*.

Els resultats obtinguts de la combinació dels 10 últims models generats per cada xarxa es mostren a la taula 4.1. L'*accuracy* d'aquesta combinació de models és del 87%. Els valors de *precision*, *recall* i *f1-score* demostren que la xarxa és capaç de diferenciar les dues classes amb prou fidelitat. La mateixa conclusió es pot extraure de la corba ROC, l'àrea sota la qual és de 0.890. La corba ROC està representada a la figura 4.6.

Taula 4.1: Resultats en el conjunt de test utilitzant xarxes neuronals. Aquestos resultats s'han obtingut a partir del model final generat per la xarxa proposada.

	precision	recall	f1-score
AD	0.78	0.77	0.77
CN	0.91	0.91	0.91
accuracy	0.87		

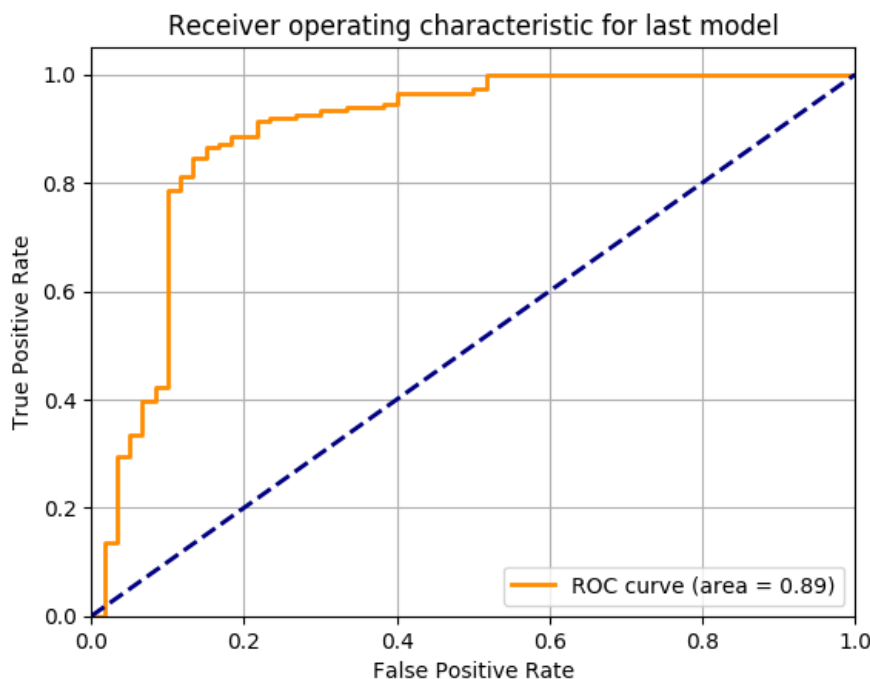


Figura 4.6: Corba ROC utilitzant les dades de volumetria i l'últim model generat per la xarxa neuronal proposada. L'àrea que s'obté amb el conjunt de test és de 0.890.

A la taula 4.2 s'indiquen els resultats produïts per la combinació dels 10 models amb major *accuracy*. Aquests resultats són lleugerament pitjors que els anteriors, tot i que l'àrea sota la corba ROC, representada a la figura 4.7, continua sent 0.890.

Taula 4.2: Resultats en el conjunt de test utilitzant xarxes neuronals. Aquests resultats s'han obtingut a partir del model amb major *accuracy* en el conjunt de validació generat per la xarxa proposada durant l'entrenament.

	precision	recall	f1-score
AD	0.76	0.63	0.69
CN	0.86	0.92	0.89
accuracy	0.84		

Finalment, els resultats obtinguts a partir de la combinació dels 10 models amb menor *loss* es poden veure a la taula 4.3. Els resultats varien poc d'aquells produïts pels models amb millor *accuracy*. No obstant, l'àrea sota la corba ROC empitjora i es queda en 0.860. La corba es pot veure a la figura 4.8.

Taula 4.3: Resultats en el conjunt de test utilitzant xarxes neuronals. Aquests resultats s'han obtingut a partir del model amb menor *loss* en el conjunt de validació generat per la xarxa proposada durant l'entrenament.

	precision	recall	f1-score
AD	0.74	0.70	0.72
CN	0.88	0.90	0.89
accuracy	0.84		

Per tant, utilitzant dades de volumetria per entrenar xarxes neuronals, els millors resultats s'han produït amb la combinació dels 10 models finals generats per cadascuna de les xarxes neuronals entrenades. Amb aquest procediment, s'ha aconseguit una *accuracy* del 87% i un *f1-score* de 0.77 per a la classe AD i 0.91 per a la classe CN. Quant a la cor-

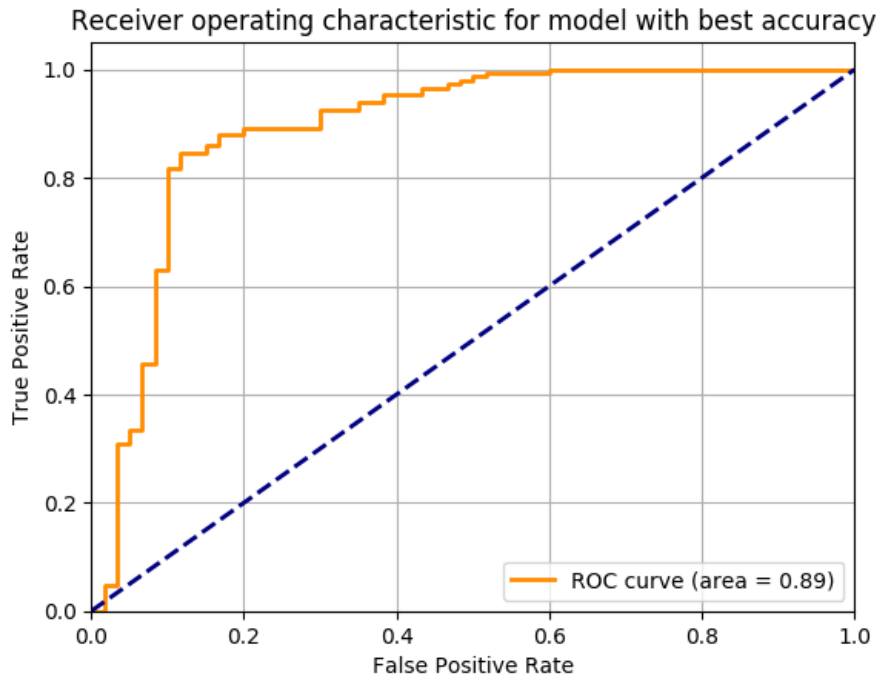


Figura 4.7: Corba ROC utilitzant les dades de volumetria i el model generat per la xarxa neuronal proposada que millor *accuracy* ha aconseguit en el conjunt de validació. L'àrea que s'obté amb el conjunt de test és de 0.890.

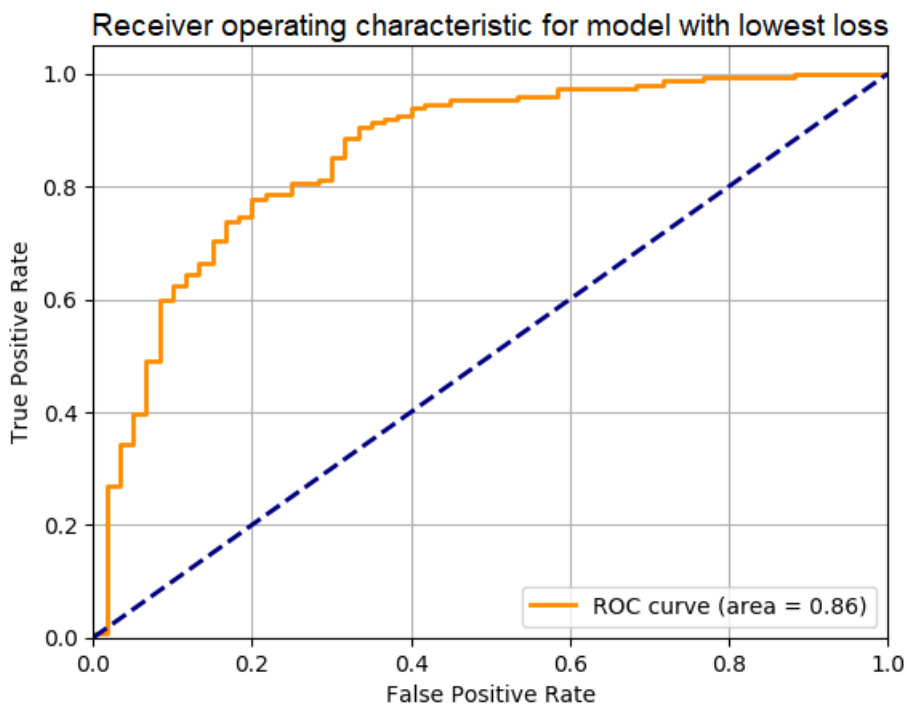


Figura 4.8: Corba ROC utilitzant les dades de volumetria i el model generat per la xarxa neuronal proposada que ha aconseguit menor *loss* en el conjunt de validació. L'àrea que s'obté amb el conjunt de test és de 0.860.

ba ROC, aquesta demostra que el model escollit és capaç de diferenciar les dues classes utilitzant els marcadors escollits utilitzant PCA.

4.4 Experiments amb mètodes de *Machine Learning*

Tal i com s'ha fet en la secció anterior, per a realitzar els experiments amb mètodes de *machine learning* utilitzant les dades de volumetria s'ha aplicat PCA per a escollir el mínim nombre de components principals de manera que es conserve el 95% de la variància. El nombre total de components principals utilitzats és 46.

4.4.1. *Random Forests*

Podem veure els resultats de validació i test a les taules 4.4 i 4.5, respectivament. Utilitzant *random forests* s'ha arribat a una *accuracy* del 85%. El millor resultat el trobem al *recall* de la classe CN en test, que és de 0.96, però el valor de *recall* per a la classe AD en test és molt baix, tan sols un 0.57. Aquests resultats ens indiquen que pràcticament cap pacient és diagnosticat falsament amb Alzheimer, però molts casos d'Alzheimer no són detectats. L'àrea sota la corba ROC en test per a la classe CN és de 0.843 i podem veure la seua representació a la figura 4.9.

Taula 4.4: Resultats en el conjunt de validació utilitzant un classificador basat en *Random Forests*.

	precision	recall	f1-score
AD	0.73	0.65	0.69
CN	0.86	0.90	0.88
accuracy	0.83		

Taula 4.5: Resultats en el conjunt de test utilitzant un classificador basat en *Random Forests*.

	precision	recall	f1-score
AD	0.85	0.57	0.68
CN	0.85	0.96	0.90
accuracy	0.85		

4.4.2. *Extremely Randomized Trees*

Com podem observar a les taules 4.6 i 4.7, el mètode basat en *extremely randomized trees* mostra encara més inclinació a classificar en CN les mostres que rep. Mentre que a la classe CN s'arriba a un *recall* de 0.97, un valor molt proper al màxim, a la classe AD aquest valor es queda en un 0.40. Tot i això, la *precision* en ambdós casos no baixa del 0.8. L'*accuracy* utilitzant aquest mètode és del 83%.

A la figura 4.10 podem veure la corba ROC, l'àrea sota la qual arriba a un 0.87 per a la classe CN en test.

4.4.3. *Bagging*

Amb *bagging* s'ha arribat a una *accuracy* del 79% i, segons podem veure a les taules 4.8 i 4.9, per a la classe CN la *precision* i *recall* –i, per tant, la *f1-score*– arriben al 0.85 però a la classe AD, aquests valors es queden en un 0.63.

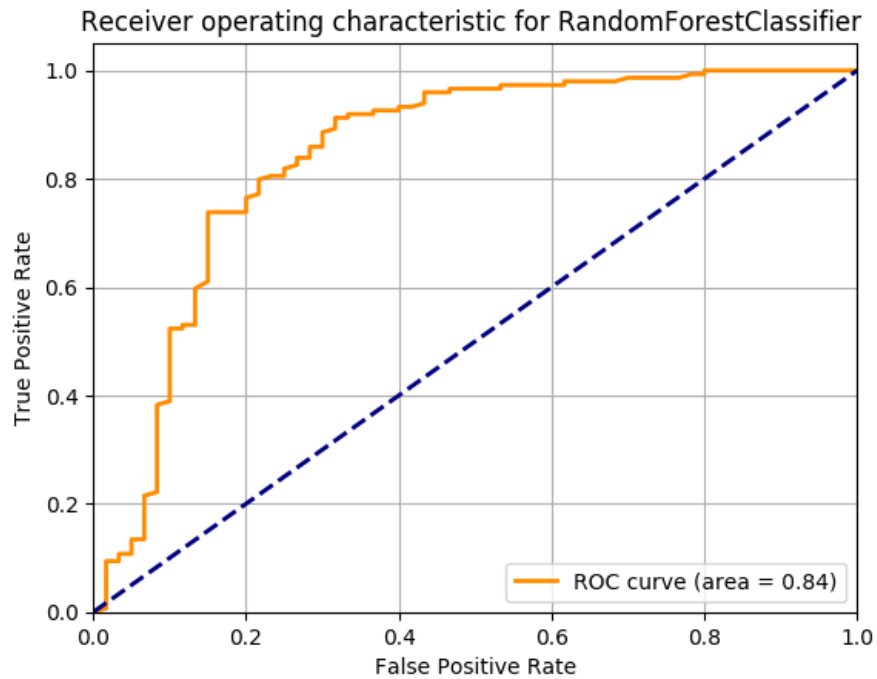


Figura 4.9: Corba ROC utilitzant un classificador basat en *Random Forests*. L'àrea que s'obté amb el conjunt de test és de 0.843.

Taula 4.6: Resultats en el conjunt de validació utilitzant un classificador basat en *Extremely Randomized Trees*.

	precision	recall	f1-score
AD	0.80	0.57	0.67
CN	0.84	0.94	0.89
accuracy			0.83

Taula 4.7: Resultats en el conjunt de test utilitzant un classificador basat en *Extremely Randomized Trees*.

	precision	recall	f1-score
AD	0.86	0.40	0.55
CN	0.80	0.97	0.88
accuracy			0.81

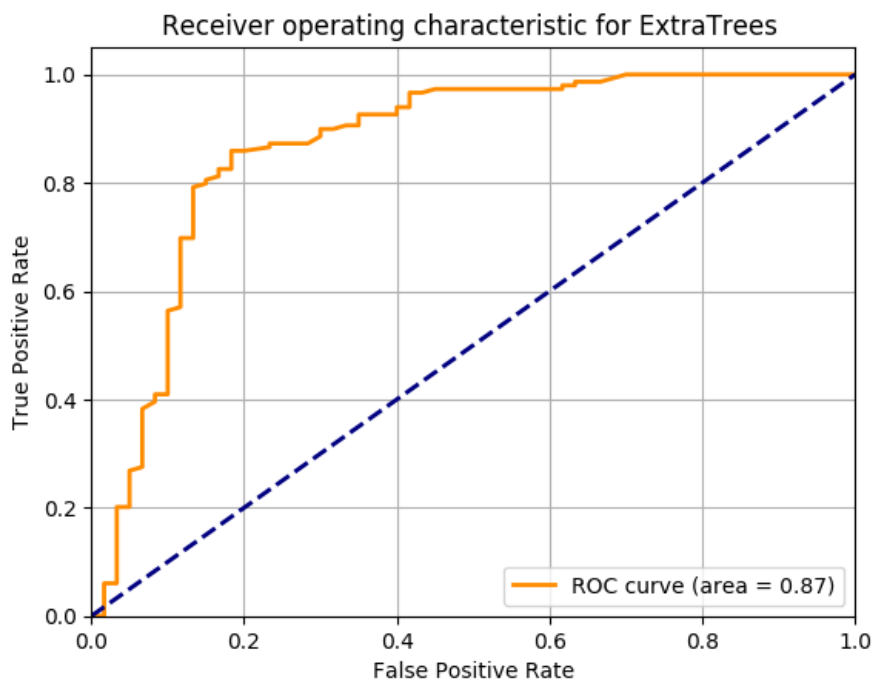


Figura 4.10: Corba ROC utilitzant un classificador basat en *Extremely Randomized Trees*. L'àrea que s'obté amb el conjunt de test és de 0.869.

Tot i que hi ha menys diferència entre els valors de *precision* i *recall* de les dues classes que en els mètodes anteriors, els resultats continuen sent bastant millorables, especialment per a la classe AD.

L'àrea sota la corba ROC en aquest cas aconsegueix un valor de 0.805. Per tant, en el cas de *bagging*, les característiques de les mostres d'una classe i de l'altra no resulten clarament determinants per a la classificació en aquestes dues classes. Es pot veure la representació de la corba a la figura 4.11.

Taula 4.8: Resultats en el conjunt de validació utilitzant un classificador basat en *Bagging*.

	precision	recall	f1-score
AD	0.69	0.65	0.67
CN	0.86	0.88	0.87
accuracy	0.81		

Taula 4.9: Resultats en el conjunt de test utilitzant un classificador basat en *Bagging*.

	precision	recall	f1-score
AD	0.63	0.63	0.63
CN	0.85	0.85	0.85
accuracy	0.79		

4.4.4. *AdaBoost*

Com podem veure a les taules 4.10 i 4.11, el mètode *AdaBoost* aconsegueix una *accuracy* del 83% en test. A més a més, s'obtenen uns valors de *precision* i *recall* per a la classe AD superiors als dels mètodes anteriors i es redueix la diferència entre els resultats per a la classe CN i la classe AD.

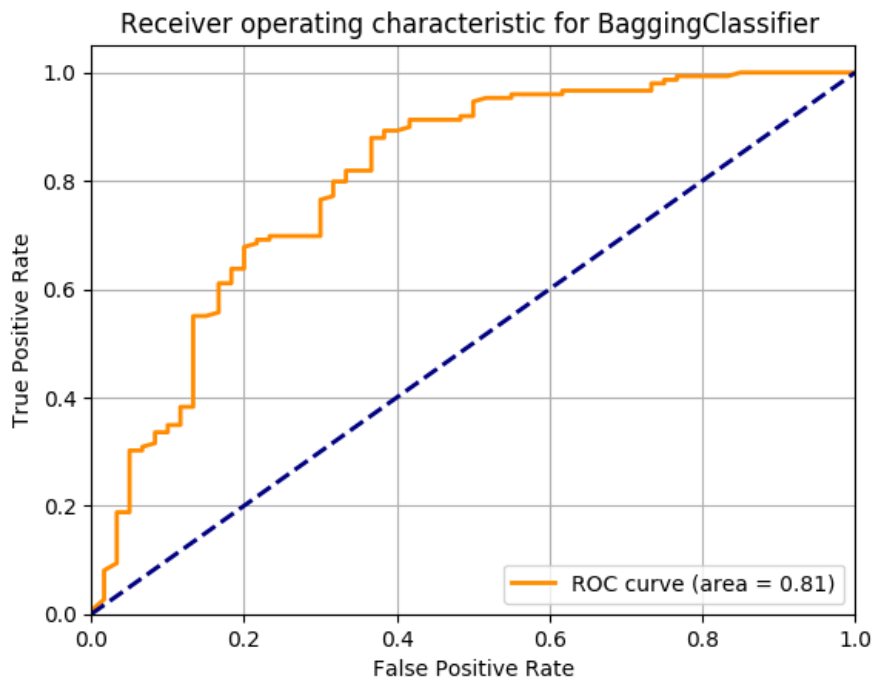


Figura 4.11: Corba ROC utilitzant un classificador basat en *Bagging*. L'àrea que s'obté amb el conjunt de test és de 0.805.

A la gràfica 4.12 podem veure la corba ROC, que forma una àrea de 0.855, la qual s'aproxima a l'àrea formada per la corba ROC utilitzant *extremely randomized trees*.

Taula 4.10: Resultats en el conjunt de validació utilitzant un classificador basat en *AdaBoost*.

	precision	recall	f1-score
AD	0.71	0.78	0.74
CN	0.90	0.87	0.89
accuracy	0.84		

Taula 4.11: Resultats en el conjunt de test utilitzant un classificador basat en *AdaBoost*.

	precision	recall	f1-score
AD	0.75	0.63	0.68
CN	0.86	0.91	0.89
accuracy	0.83		

4.4.5. Gradient Tree Boosting

A les taules 4.12 i 4.13 podem veure que l'*accuracy* obtinguda utilitzant *gradient tree boosting* es troba al voltant del 85%. Els valors de *precision*, *recall* i *f1-score* es troben al voltant del 0.7 per a la classe AD i al voltant del 0.9 per a la classe CN. Podem assumir, per tant, que tant els falsos positius com els falsos negatius són molt baixos en relació al nombre de pacients per a cada diagnòstic. L'àrea sota la corba ROC que aconseguim en test amb *gradient boosting* arriba al 0.87, i es pot veure a la figura 4.13.

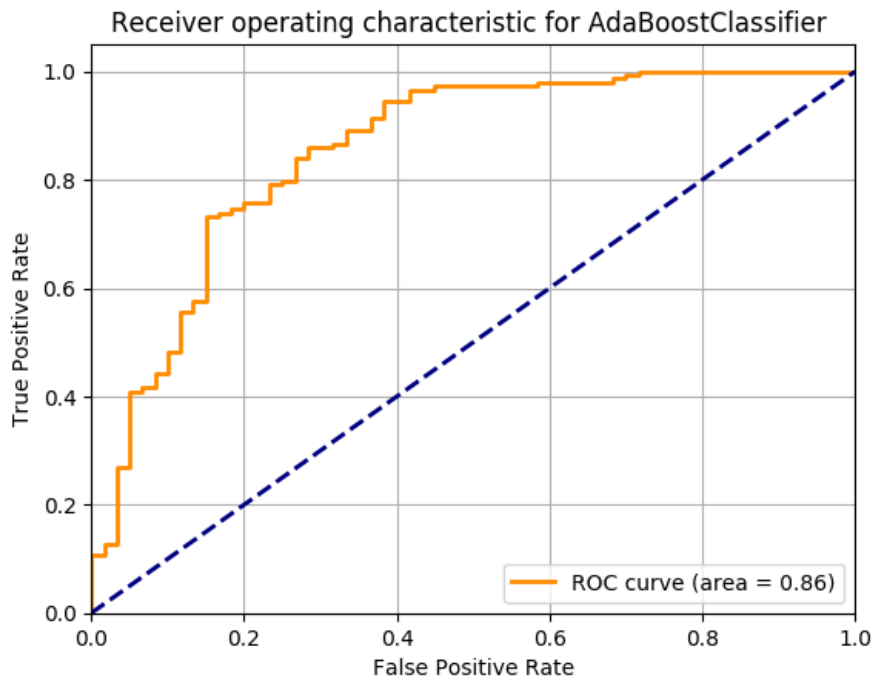


Figura 4.12: Corba ROC utilitzant un classificador basat en *AdaBoost*. L'àrea que s'obté amb el conjunt de test és de 0.855.

Taula 4.12: Resultats en el conjunt de validació utilitzant un classificador basat en *Gradient Tree Boosting*.

	precision	recall	f1-score
AD	0.73	0.71	0.72
CN	0.88	0.89	0.89
accuracy	0.84		

Taula 4.13: Resultats en el conjunt de test utilitzant un classificador basat en *Gradient Tree Boosting*.

	precision	recall	f1-score
AD	0.78	0.65	0.71
CN	0.87	0.93	0.90
accuracy	0.85		

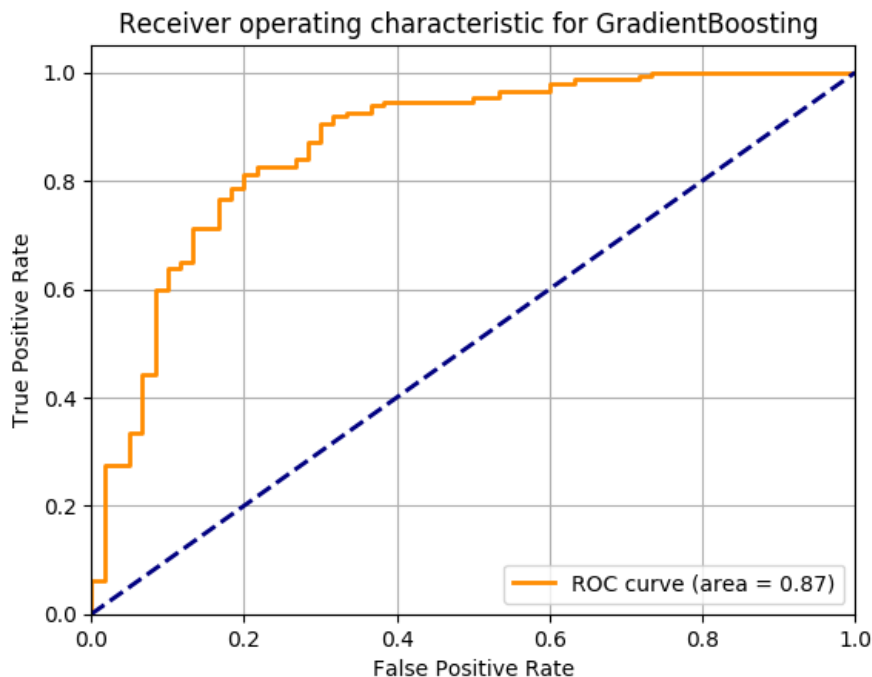


Figura 4.13: Corba ROC utilitzant un classificador basat en *Gradient Tree Boosting*. L'àrea que s'obté amb el conjunt de test és de 0.869.

4.4.6. *K-Nearest Neighbors*

Si apliquem *k-nearest neighbors* obtenim una *accuracy* de 86% en test i unes *precision* de 0.81 i 0.87, com es pot observar a les taules 4.14 i 4.15. No obstant, tot i que la *recall* augmenta fins al 0.94 per a CN, en AD baixa fins al 0.65.

La figura 4.14 il·lustra la corba ROC per a KNN amb una àrea igual a 0.865.

Taula 4.14: Resultats en el conjunt de validació utilitzant un classificador basat en *K-Nearest Neighbors*.

	precision	recall	f1-score
AD	0.76	0.71	0.74
CN	0.89	0.91	0.90
accuracy	0.85		

Taula 4.15: Resultats en el conjunt de test utilitzant un classificador basat en *K-Nearest Neighbors*.

	precision	recall	f1-score
AD	0.81	0.65	0.72
CN	0.87	0.94	0.90
accuracy	0.86		

4.4.7. *Neighborhood Components Analysis*

En combinar *k-nearest neighbors* amb *neighborhood components analysis* l'*accuracy* no millora i la *precision* baixa. No obstant, tal i com podem observar a les taules 4.16 i 4.17, els valors de *recall* i *f1-score* es mantenen més o menys constants, a l'igual que la corba ROC, la qual es mostra a la figura 4.15.

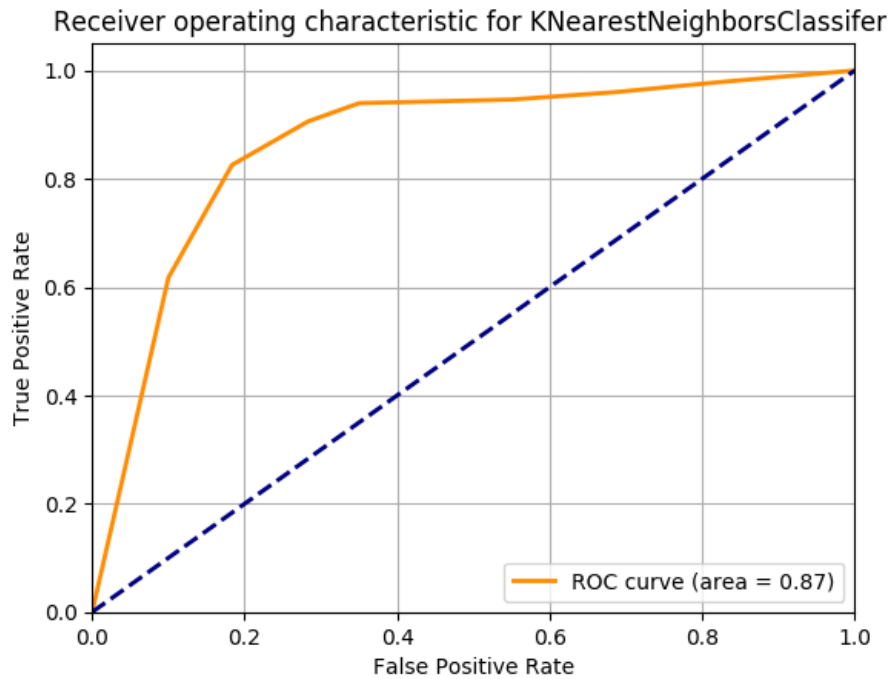


Figura 4.14: Corba ROC utilitzant un classificador basat en *K Nearest-Neighbors*. L'àrea que s'obté amb el conjunt de test és de 0.865.

Taula 4.16: Resultats en el conjunt de validació utilitzant un classificador basat en *K-Nearest Neighbors* i *Neighborhood Components Analysis*.

	precision	recall	f1-score
AD	0.69	0.67	0.68
CN	0.86	0.88	0.87
accuracy	0.81		

Taula 4.17: Resultats en el conjunt de test utilitzant un classificador basat en *K-Nearest Neighbors* i *Neighborhood Components Analysis*.

	precision	recall	f1-score
AD	0.76	0.68	0.72
CN	0.88	0.91	0.89
accuracy	0.85		

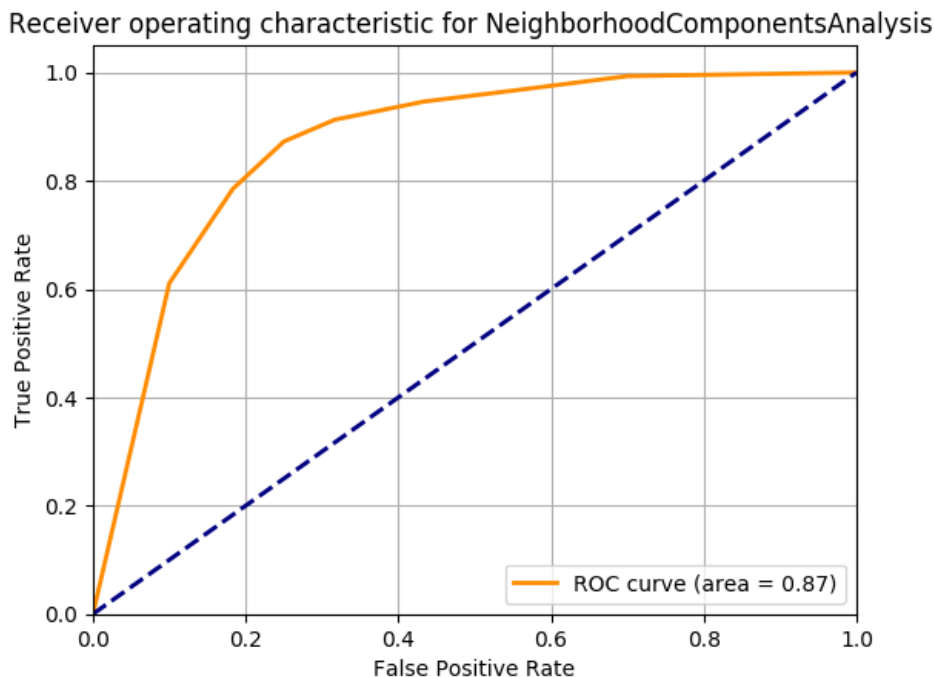


Figura 4.15: Corba ROC utilitzant un classificador basat en *K-Nearest Neighbors* i *Neighborhood Components Analysis*. L'àrea que s'obté amb el conjunt de test és de 0.869.

4.4.8. SVM

Amb la classificació per vectors de suport hem provat a utilitzar dos kernels diferents: *poly* i *rbf*.

Els resultats que hem obtingut amb el kernel *poly* es mostren a les taules 4.18 i 4.19. Podem veure que hem aconseguit una *accuracy* de 88% en el conjunt de test. A més, el resultat més alt l'hem aconseguit en *recall* per a la classe CN, un 0.93; el resultat amb menys diferència entre les classes AD i CN és la *precision*, on AD té un 0.81 i CN un 0.90. Aquests resultats són positius i indiquen que el nombre de mostres incorrectament classificades és baix per a les dues classes.

Podem veure la representació de la corba ROC, amb una àrea de 0.908, a la figura 4.16. Aquesta àrea és superior a l'aconseguida amb els experiments anteriors.

Taula 4.18: Resultats en el conjunt de validació utilitzant un classificador basat en SVM amb kernel *poly*.

	precision	recall	f1-score
AD	0.75	0.75	0.75
CN	0.90	0.90	0.90
accuracy	0.85		

Taula 4.19: Resultats en el conjunt de test utilitzant un classificador basat en SVM amb kernel *poly*.

	precision	recall	f1-score
AD	0.81	0.73	0.77
CN	0.90	0.93	0.91
accuracy	0.88		

Receiver operating characteristic for SVC C=1.0 coef0=1.0 degree=3 kernel=poly

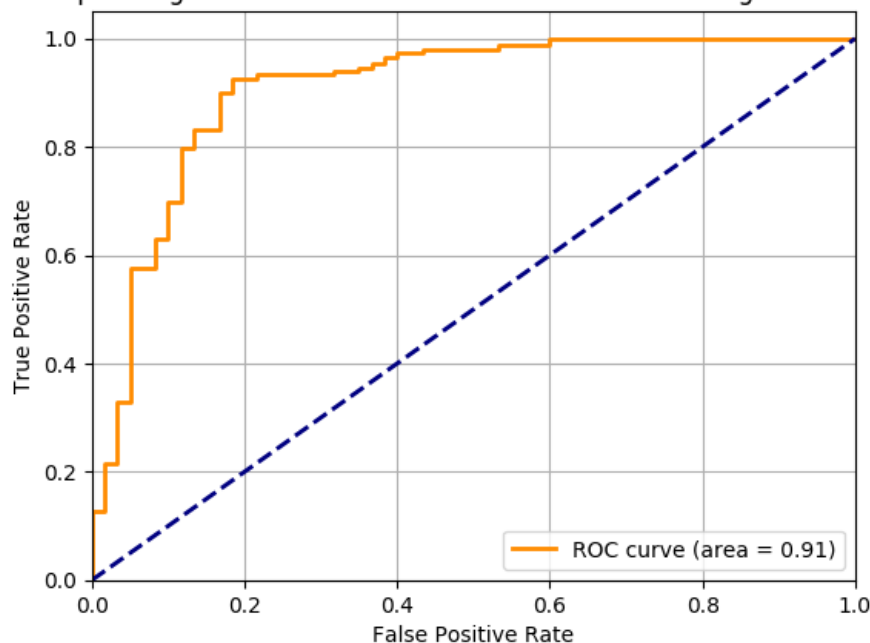


Figura 4.16: Corba ROC utilitzant un classificador basat en SVM amb kernel *poly*. L'àrea que s'obté amb el conjunt de test és de 0.908.

Si, en canvi, utilitzem un kernel *rbf*, no notarem gran diferència, ja que l'*accuracy*, com podem veure a les taules 4.20 i 4.21, és de 83% en development i de 86% en test. Tant la *precision* com la *recall* i la *f1-score* empitjoren lleugerament, sense significar una canvi notable en els resultats, com també ho fa l'àrea de la corba ROC, la qual es mostra a la figura 4.17.

Taula 4.20: Resultats en el conjunt de validació utilitzant un classificador basat en SVM amb kernel *rbf*.

	precision	recall	f1-score
AD	0.74	0.67	0.70
CN	0.87	0.90	0.88
accuracy	0.83		

Taula 4.21: Resultats en el conjunt de test utilitzant un classificador basat en SVM amb kernel *rbf*.

	precision	recall	f1-score
AD	0.77	0.73	0.75
CN	0.89	0.91	0.90
accuracy	0.86		

La funció NuSVC és similar a la funció SVC però utilitzant un paràmetre per a controlar el nombre de vectors de suport. Hem provat dos versions variant el valor del paràmetre *nu*. Aquest paràmetre marca un límit superior de la fracció dels errors d'entrenament i un límit inferior de la fracció de vectors de suport. En ambdós casos hem utilitzat el kernel *rbf*.

En la versió on *nu* és igual a 0.01, s'ha obtingut una *accuracy* de 86% en test. A més, la *f1-score* arriba al 0.90 per a CN i al 0.75 per a AD en test. Els resultats es mostren a les taules 4.22 i 4.23.

Receiver operating characteristic for SVC C=1.0 coef0=1.0 degree=3 kernel=rbf

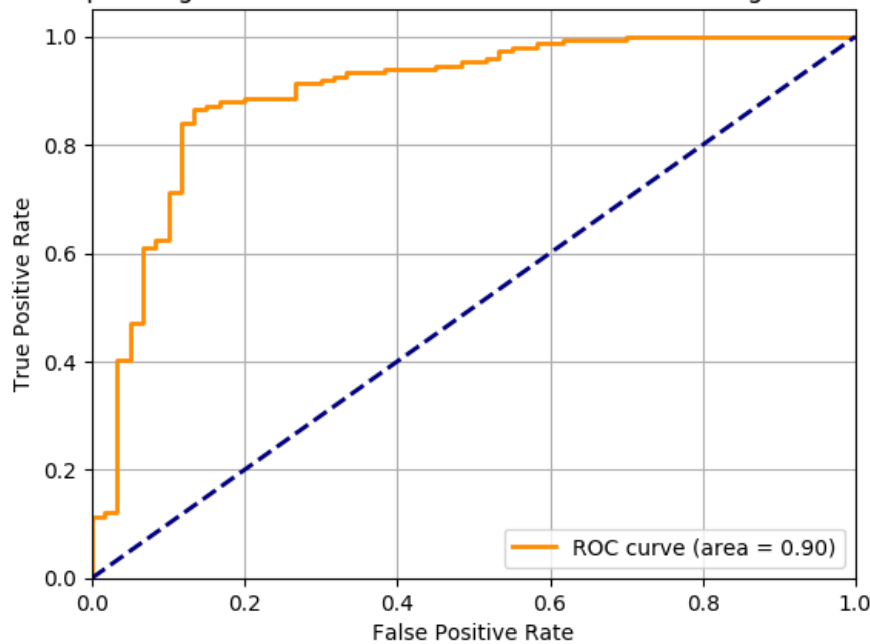


Figura 4.17: Corba ROC utilitzant un classificador basat en SVM amb kernel *rbf*. L'àrea que s'obté amb el conjunt de test és de 0.897.

El valor de l'àrea sota la corba ROC, representada en la figura 4.18, en aquest cas arriba a un 0.896.

Taula 4.22: Resultats en el conjunt de validació utilitzant un classificador basat en SVM amb $\nu=0.01$.

	precision	recall	f1-score
AD	0.74	0.67	0.70
CN	0.87	0.90	0.88
accuracy	0.83		

Taula 4.23: Resultats en el conjunt de test utilitzant un classificador basat en SVM amb $\nu=0.01$.

	precision	recall	f1-score
AD	0.76	0.73	0.75
CN	0.89	0.91	0.90
accuracy	0.86		

Si utilitzem la funció NuSVC amb $\nu = 0.1$, els resultats no canvien gens en comparació amb l'ús de $\nu = 0.01$ i amb la funció SVC amb el mateix kernel, com podem veure a les taules 4.24 i 4.25 i a la gràfica 4.19, la qual cosa indica que canviar el nombre mínim de vectors de suport no afecta a la classificació amb vectors de suport i kernel *poly* o que, potser, es necessita un valor més alt de ν per a millorar els resultats.

Per tant, entre les tècniques basades en SVM, la que millors resultats produeix és el classificador basat en SVM amb kernel *poly*, sense especificar prèviament el nombre de vectors de suport.

Receiver operating characteristic for NuSVC nu=0.1 coef0=1.0 degree=3 kernel=rbf

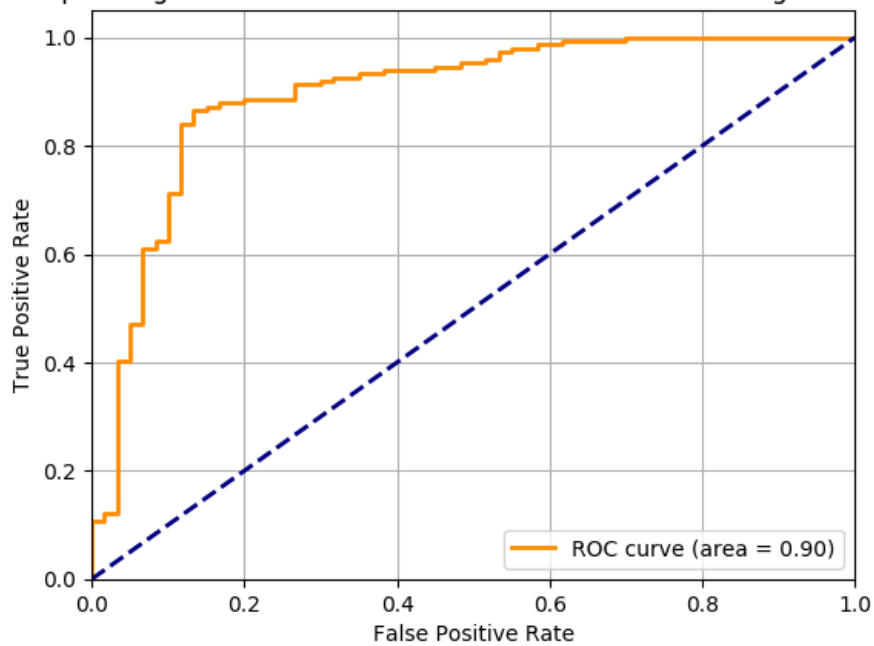


Figura 4.18: Corba ROC utilitzant un classificador basat en SVM amb $\nu=0.01$. L'àrea que s'obté amb el conjunt de test és de 0.900.

Taula 4.24: Resultats en el conjunt de validació utilitzant un classificador basat en SVM amb $\nu=0.1$.

	precision	recall	f1-score
AD	0.74	0.67	0.70
CN	0.87	0.90	0.88
accuracy	0.83		

Taula 4.25: Resultats en el conjunt de test utilitzant un classificador basat en SVM amb $\nu=0.1$.

	precision	recall	f1-score
AD	0.76	0.73	0.75
CN	0.89	0.91	0.90
accuracy	0.86		

Receiver operating characteristic for NuSVC nu=0.01 coef0=1.0 degree=3 kernel=rbf

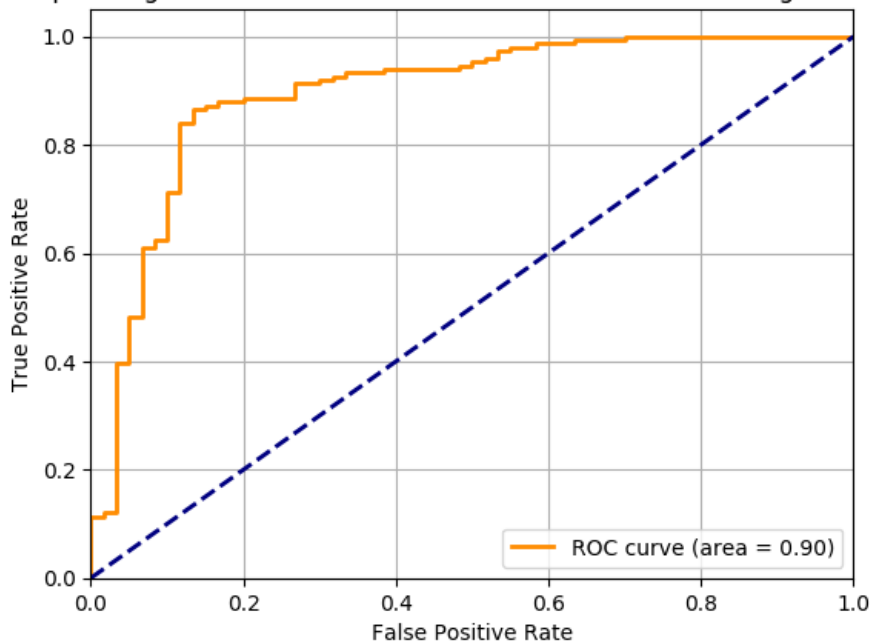


Figura 4.19: Corba ROC utilitzant un classificador basat en SVM amb $nu=0.1$. L'àrea que s'obté amb el conjunt de test és de 0.896.

4.4.9. Gaussian Naive Bayes

Amb l'ús del mètode *gaussian naive Bayes* s'aconsegueix una *accuracy* del 80% en test. El millor resultat el trobem en el *recall* de la classe CN, amb 0.91; però el *recall* de la classe AD es queda en 0.70. Les taules 4.26 i 4.27 mostren aquests resultats.

Quant a l'àrea sota la corba ROC, representada a la figura 4.20, aquesta té un valor de 0.815.

Taula 4.26: Resultats en el conjunt de validació utilitzant un classificador basat en *Gaussian Naive Bayes*.

	precision	recall	f1-score
AD	0.70	0.70	0.70
CN	0.88	0.88	0.88
accuracy			0.82

Taula 4.27: Resultats en el conjunt de test utilitzant un classificador basat en *Gaussian Naive Bayes*.

	precision	recall	f1-score
AD	0.70	0.70	0.70
CN	0.83	0.91	0.87
accuracy			0.80

4.4.10. Gaussian Process

Utilitzant *gaussian process* s'aconsegueix una *accuracy* del 79%, amb uns resultats pobres per a la classe AD en *recall*, amb un valor de 0.57, i també en *precision*, amb un 0.65. En la classe CN el valor de *precision* és 0.83 i el de *recall* és 0.88. Els resultats es mostren a les taules 4.28 i 4.29.

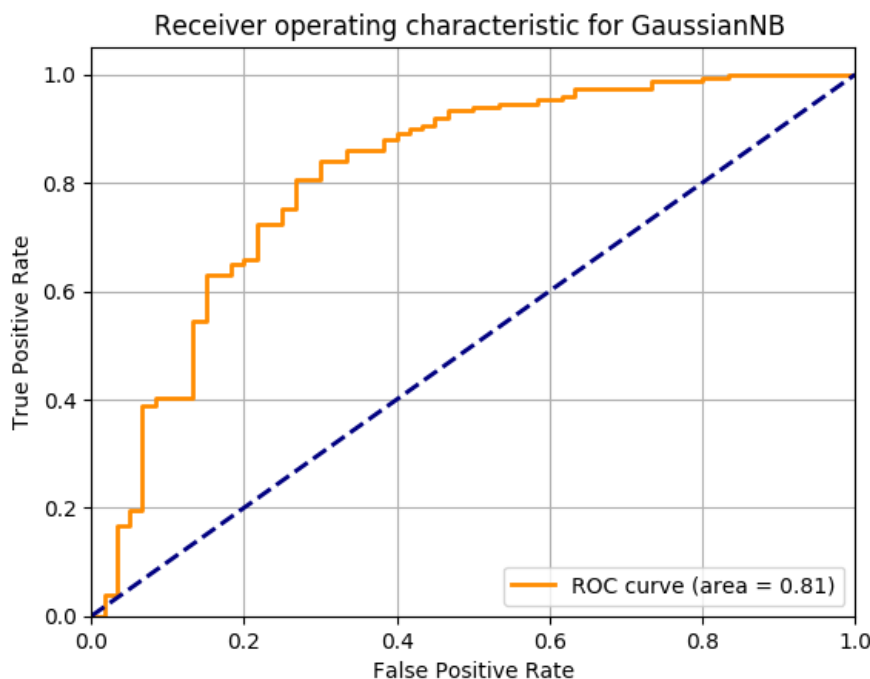


Figura 4.20: Corba ROC utilitzant un classificador basat en *Gaussian Naive Bayes*. L'àrea que s'obté amb el conjunt de test és de 0.815.

La corba ROC resultant d'aplicar *gaussian process* forma una àrea de 0.786, com s'observa a la figura 4.21.

Taula 4.28: Resultats en el conjunt de validació utilitzant un classificador basat en *Gaussian Process*.

	precision	recall	f1-score
AD	0.61	0.67	0.64
CN	0.86	0.82	0.84
accuracy	0.78		

Taula 4.29: Resultats en el conjunt de test utilitzant un classificador basat en *Gaussian Process*.

	precision	recall	f1-score
AD	0.65	0.57	0.61
CN	0.83	0.88	0.86
accuracy	0.79		

4.4.11. *Linear Discriminant Analysis* i *Quadratic Discriminant Analysis*

Si comparem els resultats de l'anàlisi discriminant lineal –taules 4.30 i 4.31– i l'anàlisi discriminant quadràtic –taules 4.32 i 4.33–, advertim uns millors resultats a l'anàlisi lineal. Amb aquest s'arriba a una *accuracy* del 85% mentre que el quadràtic es queda en un 80%. Els resultats de *f1-score* també són millors a l'anàlisi lineal, on el valor per a la classe AD és 0.74 i el valor per a la classe CN és 0.90. A l'anàlisi quadràtic el *f1-score* es queda en 0.63 per a la classe AD i en 0.86 per a la classe CN.

La diferència en l'àrea sota la corba ROC també és considerable: a l'anàlisi lineal arriba al 0.881, però a l'anàlisi quadràtic es queda per baix del 0.8 amb sols un 0.756. Aquestes corbes s'il·lustren a les figures 4.22 i 4.23.

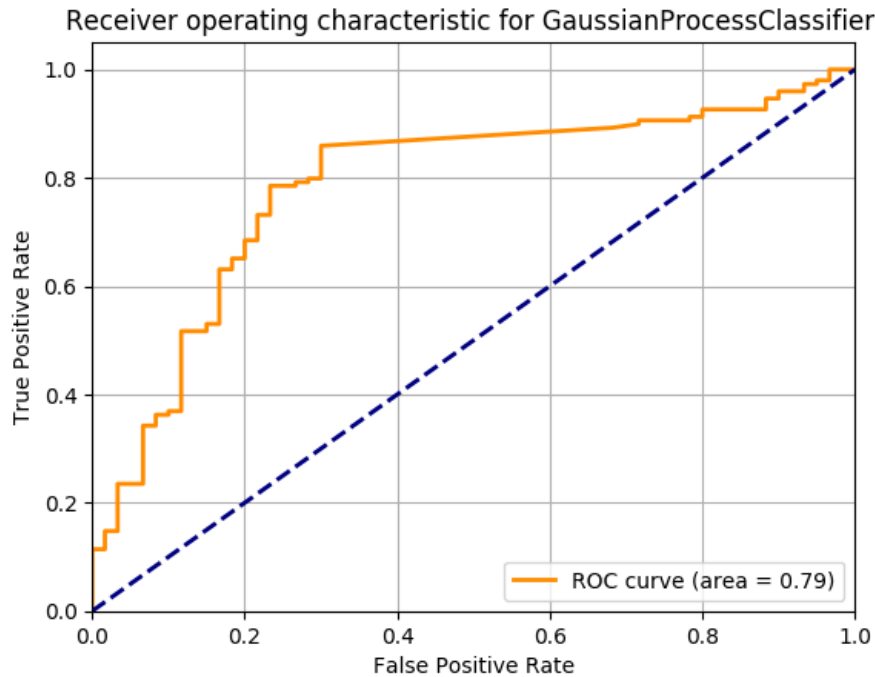


Figura 4.21: Corba ROC utilitzant un classificador basat en *Gaussian Process*. L'àrea que s'obté amb el conjunt de test és de 0.786.

Taula 4.30: Resultats en el conjunt de validació utilitzant un classificador basat en *Linear Discriminant Analysis*.

	precision	recall	f1-score
AD	0.73	0.71	0.72
CN	0.88	0.89	0.89
accuracy	0.84		

Taula 4.31: Resultats en el conjunt de test utilitzant un classificador basat en *Linear Discriminant Analysis*.

	precision	recall	f1-score
AD	0.75	0.72	0.74
CN	0.89	0.91	0.90
accuracy	0.85		

Taula 4.32: Resultats en el conjunt de validació utilitzant un classificador basat en *Quadratic Discriminant Analysis*.

	precision	recall	f1-score
AD	0.67	0.67	0.67
CN	0.86	0.86	0.86
accuracy	0.81		

Taula 4.33: Resultats en el conjunt de test utilitzant un classificador basat en *Quadratic Discriminant Analysis*.

	precision	recall	f1-score
AD	0.67	0.58	0.63
CN	0.84	0.89	0.86
accuracy	0.80		

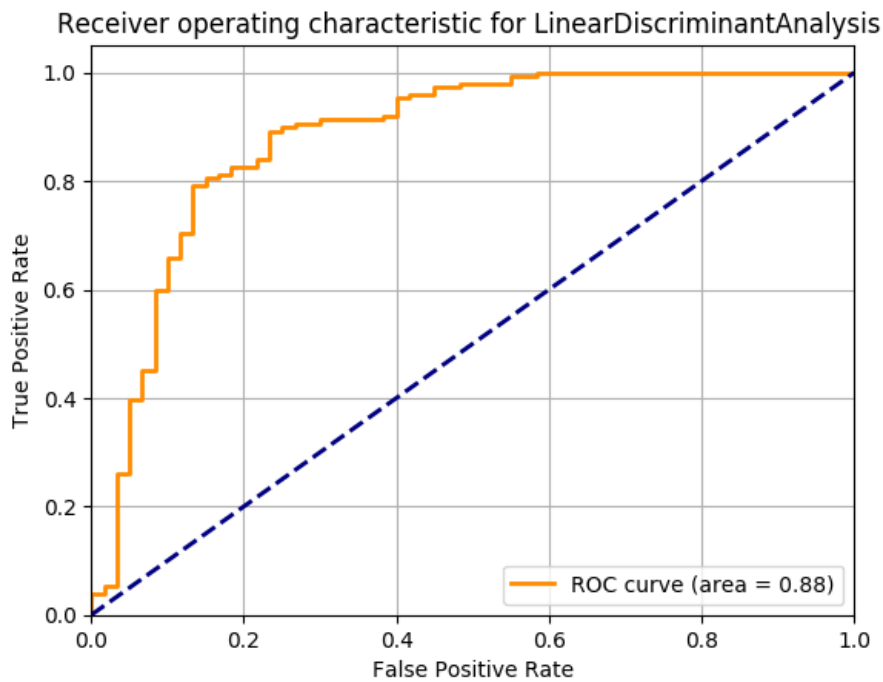


Figura 4.22: Corba ROC utilitzant un classificador basat en *Linear Discriminant Analysis*. L'àrea que s'obté amb el conjunt de test és de 0.881.

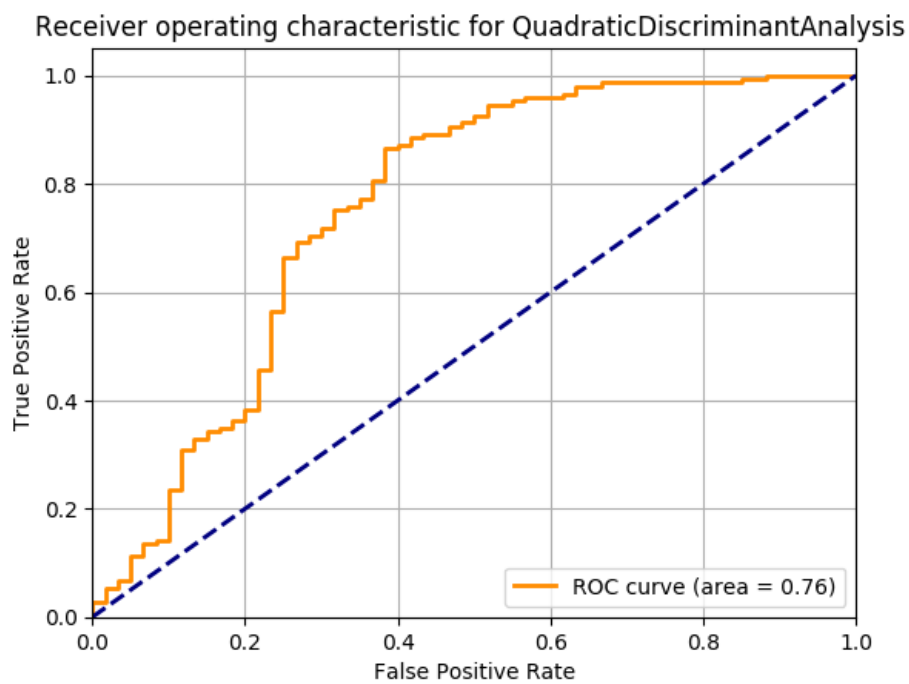


Figura 4.23: Corba ROC utilitzant un classificador basat en *Quadratic Discriminant Analysis*. L'àrea que s'obté amb el conjunt de test és de 0.756.

4.4.12. Bayesian Gaussian Mixture

Amb les mixtures de gaussianes s'han aconseguit resultats molt positius en *recall* per a la classe CN amb un 0.95, que contrasten amb el valor de *recall* per a la classe AD, que té un 0.52. La diferència no és tan gran en *precision*, per la qual cosa s'arriba a un 0.63 i un 0.88 en *f1-score* per a les classes AD i CN, respectivament. L'*accuracy* en aplicar aquest mètode és del 82%. Els resultats es mostren a les taules 4.34 i 4.35.

L'àrea sota la corba ROC és de 0.824, com es pot veure a la figura 4.24.

Taula 4.34: Resultats en el conjunt de validació utilitzant un classificador basat en *Bayesian Gaussian Mixture*.

	precision	recall	f1-score
AD	0.76	0.62	0.68
CN	0.85	0.92	0.89
accuracy			0.83

Taula 4.35: Resultats en el conjunt de test utilitzant un classificador basat en *Bayesian Gaussian Mixture*.

	precision	recall	f1-score
AD	0.79	0.52	0.63
CN	0.83	0.95	0.88
accuracy			0.82

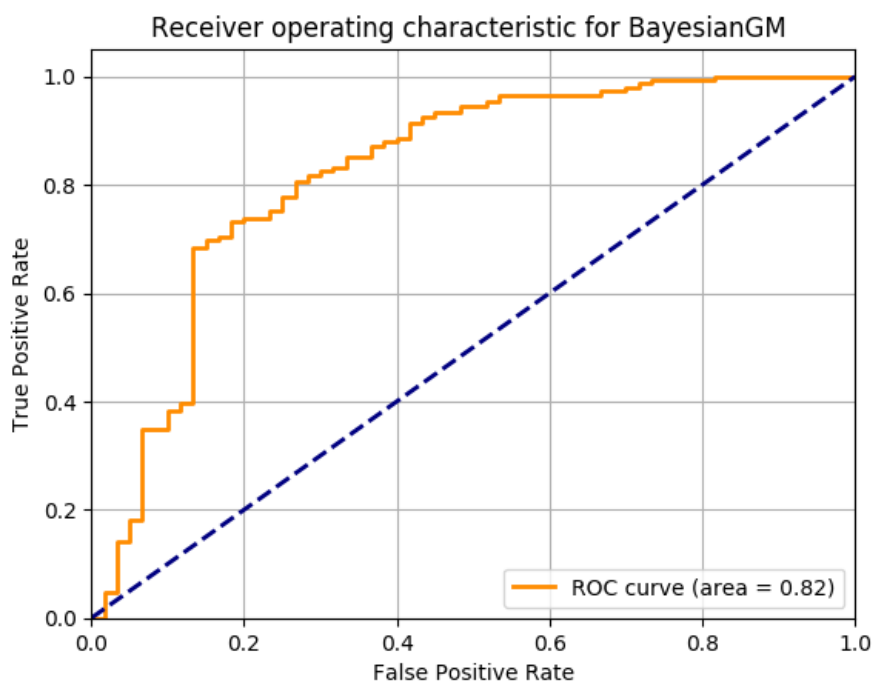


Figura 4.24: Corba ROC utilitzant un classificador basat en *Bayesian Gaussian Mixture*. L'àrea que s'obté amb el conjunt de test és de 0.824.

CAPÍTOL 5

Conclusions

Amb aquest projecte es demostra la rellevància que tenen les dades de volumetria que poden obtindre's del cervell per a diagnosticar la malaltia d'Alzhèimer. S'obtenen uns resultats positius tant amb l'aplicació de xarxes neuronals com amb l'ús de mètodes de *machine learning*, tot i que no hem aconseguit arribar als resultats publicats als articles de l'estat de l'art ja que no hem pogut replicar els experiments de forma exacta. A banda dels bons resultats obtinguts, l'ús de dades de volumetria presenta l'avantatge d'uns menors costos temporals i de computació, a excepció del preprocessament amb FreeSurfer, per a implementar classificadors en comparació amb els recursos necessaris si s'utilitzen les imatges.

Els millors resultats utilitzant mètodes de *machine learning* s'han obtingut gràcies a aplicar *support vector machines*, tot i que els resultats de *linear discriminant analysis* no es queden molt per darrere. Aplicant SVM amb kernel *poly* s'ha aconseguit una *accuracy* del 88% en la classificació entre les classes AD i CN. Els valors de *precision* obtinguts són 0.81 per a la classe AD i 0.90 per a la classe CN. En *recall* els resultats són 0.73 per a AD i 0.93 per a CN. El *f1-score* és 0.77 per a AD i 0.91 per a CN. Aquestos resultats demostren que el classificador implementat tendeix a detectar millor els casos de cervells sans, tot i que també descobreix la major part dels casos d'Alzhèimer. L'àrea sota la corba ROC amb SVM utilitzant kernel *poly* és de 0.908, la qual cosa indica que els components escollits per a realitzar la classificació de les mostres resulten considerablement distintius per a l'aplicació d'SVM.

Quant a l'ús de xarxes neuronals, s'ha pogut comprovar que aquestes són capaces d'igualar i, en molts casos, millorar els resultats dels mètodes clàssics de *machine learning*. En els nostres experiments, l'*accuracy* ha sigut del 87% i els valors de *f1-score* han sigut 0.77 per a la classe AD i 0.91 per a la classe CN. L'àrea sota la corba ROC ha sigut de 0.890. A la vista d'aquests resultats, podem concloure que la investigació de xarxes neuronals aplicades a dades de volumetria extretes d'imatges de cervell és un camí a seguir prometedor.

L'objectiu futur principal consisteix en combinar les dades de volumetria amb imatges per a utilitzar-les en xarxes neuronals. És interessant, en especial, investigar amb imatges de tipus MRI solament, ja que s'obtenen a partir d'un procediment no invasiu, al contrari que les imatges PET, les quals són molt utilitzades als experiments de l'estat de l'art.

Ens haurien sigut de gran ajuda unes explicacions més detallades i concretes dels passos seguits a les investigacions relatades a l'estat de l'art per a poder replicar-les i, a més, comprovar la seua veracitat.

Bibliografía

- [1] F Bermejo-Pareja, J Benito-León, S Vega, MJ Medrano, GC Román, Neurological Disorders in Central Spain (NEDICES) Study Group, et al. Incidence and subtypes of dementia in three elderly populations of central Spain. *Journal of the neurological sciences*, 264(1-2):63–72, 2008.
- [2] World Health Organization and others. First WHO ministerial conference on global action against dementia: meeting report, WHO Headquarters, Geneva, Switzerland, 16-17 March 2015. *World Health Organization*, 2015.
- [3] Alzheimer’s Association. Ayuda para la enfermedad de Alzheimer y la demencia | España. <https://www.alz.org/es/>, 1980.
- [4] Manhua Liu, Danni Cheng, Kundong Wang, Yaping Wang, Alzheimer’s Disease Neuroimaging Initiative, et al. Multi-modality cascaded convolutional neural networks for Alzheimer’s disease diagnosis. *Neuroinformatics*, pages 1–14, 2018.
- [5] Hongfei Wang, Yanyan Shen, Shuqiang Wang, Tengfei Xiao, Liming Deng, Xiangyu Wang, and Xinyan Zhao. Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer’s disease. *Neurocomputing*, 333:145–156, 2019.
- [6] Karim Aderghal, Manuel Boissenin, Jenny Benois-Pineau, Gwenaëlle Catheline, and Karim Afdel. Classification of sMRI for AD Diagnosis with Convolutional Neuronal Networks: A Pilot 2-D+ ϵ Study on ADNI. In *International Conference on Multimedia Modeling*, pages 690–701. Springer, 2017.
- [7] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, Alzheimer’s Disease Neuroimaging Initiative, et al. Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database. *neuroimage*, 56(2):766–781, 2011.
- [8] Stavros I Dimitriadis, Dimitris Liparas, Magda N Tsolaki, Alzheimer’s Disease Neuroimaging Initiative, et al. Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healthy elderly, MCI, cMCI and alzheimer’s disease patients: From the alzheimer’s disease neuroimaging initiative (ADNI) database. *Journal of neuroscience methods*, 302:14–23, 2018.
- [9] Daoqiang Zhang, Yaping Wang, Luping Zhou, Hong Yuan, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Multimodal classification of alzheimer’s disease and mild cognitive impairment. *Neuroimage*, 55(3):856–867, 2011.

- [10] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems. <http://tensorflow.org/>, 2015. Software available from tensorflow.org.
- [11] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.
- [12] François Chollet et al. Keras. <https://keras.io>, 2015.
- [13] Mark W Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, and Stephen M Smith. Bayesian analysis of neuroimaging data in fsl. *Neuroimage*, 45(1):S173–S186, 2009.
- [14] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.
- [15] Stephen M Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155, 2002.
- [16] Mark Jenkinson, Mickael Pechaud, Stephen Smith, et al. BET2: MR-based estimation of brain, skull and scalp surfaces. In *Eleventh annual meeting of the organization for human brain mapping*, volume 17, page 167. Toronto., 2005.
- [17] Valeriu Popescu, Marco Battaglini, WS Hoogstrate, Sander CJ Verfaillie, IC Sluimer, Ronald A van Schijndel, Bob W van Dijk, Keith S Cover, Dirk L Knol, Mark Jenkinson, et al. Optimizing parameter choice for fsl-brain extraction tool (bet) on 3d t1 images in multiple sclerosis. *Neuroimage*, 61(4):1484–1494, 2012.
- [18] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.
- [19] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2):143–156, 2001.
- [20] Douglas N Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72, 2009.
- [21] Martin Reuter, Nicholas J Schmansky, H Diana Rosas, and Bruce Fischl. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4):1402–1418, 2012.
- [22] Bruce Fischl, David H. Salat, André J.W. van der Kouwe, Nikos Makris, Florent Ségonne, Brian T. Quinn, and Anders M. Dale. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23(Supplement 1):S69 – S84, 2004. Mathematics in Brain Imaging.

-
- [23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [24] Michael W. Weiner et al. The Alzheimer’s Disease Neuroimaging Initiative (ADNI). <http://adni.loni.usc.edu>, 2004.
- [25] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehericy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, Alzheimer’s Disease Neuroimaging Initiative, et al. Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database. *neuroimage*, 56(2):766–781, 2011.

APÈNDIX A

Estadístiques del conjunt de dades ADNI

En aquest apartat es mostra la informació detallada sobre les imatges del conjunt ADNI. A les taules [A.1](#) i [A.2](#) es fa el recompte de sessions i subjectes en rangs d'edat de cinc anys tenint en compte l'etiqueta assignada a cadascun i el sexe del pacient. A la taula [A.3](#) es mostra el resultat de l'estudi de les mides de les imatges 2D del conjunt. I a les figures [A.1](#) i [A.2](#) es mostren dos histogrames sobre el valor més alt d'intensitat de cada imatge; a la segona d'aquestes figures s'han eliminat els valors alts i poc freqüents per a veure les dades de forma més informativa.

Taula A.1: Recompte de sessions per edat, sexe i etiqueta

edat	CN		EMCI		MCI		LMCI		AD		Total edat
	F	M	F	M	F	M	F	M	F	M	
51 a 55	0	0	0	0	0	0	1	0	0	0	1
56 a 60	8	2	11	7	11	4	16	3	9	4	75
61 a 65	18	13	51	47	18	26	19	16	10	12	230
66 a 70	48	45	70	58	31	42	27	12	13	17	363
71 a 75	121	97	47	79	56	81	37	34	49	32	633
76 a 80	121	122	33	63	65	112	20	28	22	52	638
81 a 85	53	77	29	21	56	91	15	19	24	38	423
86 a 90	18	41	2	6	15	56	6	4	23	24	195
91 a 95	4	3	1	0	0	5	1	0	3	0	17
96 a 100	1	0	0	0	0	0	0	0	0	0	1
Total sexe	392	400	244	281	252	417	142	116	153	179	2576
Total etiqueta	792		525		669		258		332		

Taula A.2: Recompte de subjectes per edat, sexe i etiqueta

edat	CN		EMCI		MCI		LMCI		AD		Total edat
	F	M	F	M	F	M	F	M	F	M	
51 a 55	0	0	0	0	0	0	1	0	0	0	1
56 a 60	5	2	8	4	6	3	6	2	5	2	43
61 a 65	14	7	27	22	12	15	9	9	6	7	128
66 a 70	33	29	40	31	18	20	14	9	6	11	211
71 a 75	69	53	26	40	31	46	20	20	29	24	358
76 a 80	65	67	17	37	36	59	15	16	14	30	356
81 a 85	34	43	17	16	30	44	12	12	15	25	248
86 a 90	13	24	2	5	10	29	3	3	13	15	117
91 a 95	2	3	1	0	0	5	1	0	1	0	13
96 a 100	1	0	0	0	0	0	0	0	0	0	1
Total sexe	236	228	138	155	143	221	81	71	89	114	1476
Total etiqueta	464		293		364		152		203		

Taula A.3: Recompte de sessions per mida d'imatge 2D

Mida	Sessions
176x240	1078
160x192	682
170x256	590
160x240	190
208x240	46
211x256	24
184x256	19
176x256	8
230x230	6
230x240	5
170x288	3
176x248	3
140x256	2
150x256	1
512x174	1
160x256	1

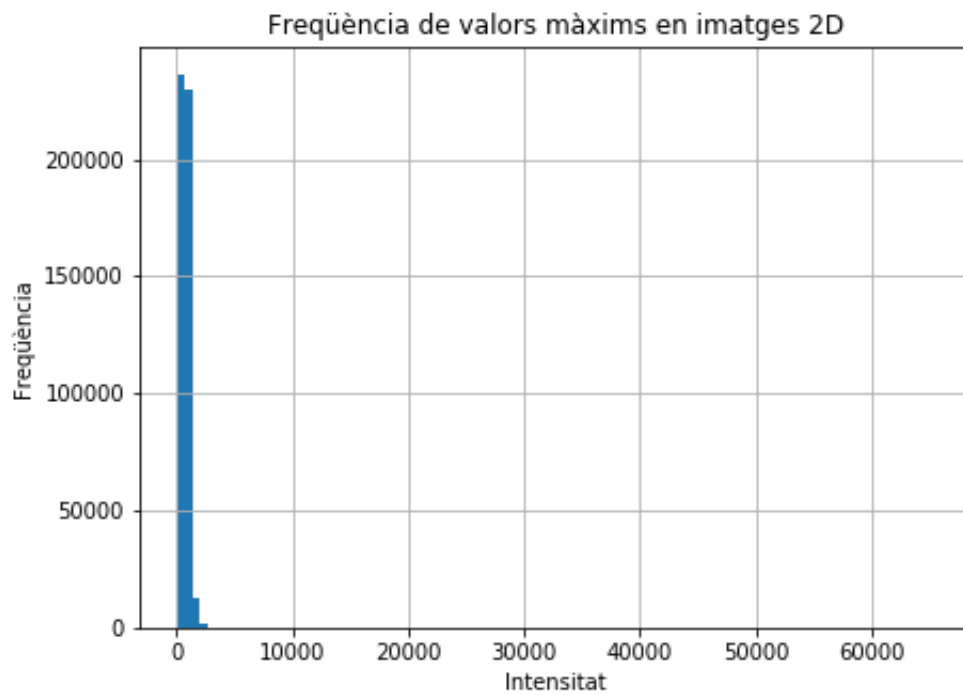


Figura A.1: Freqüència dels valors màxims d'intensitat de cadascuna de les imatges 2D del conjunt ADNI.

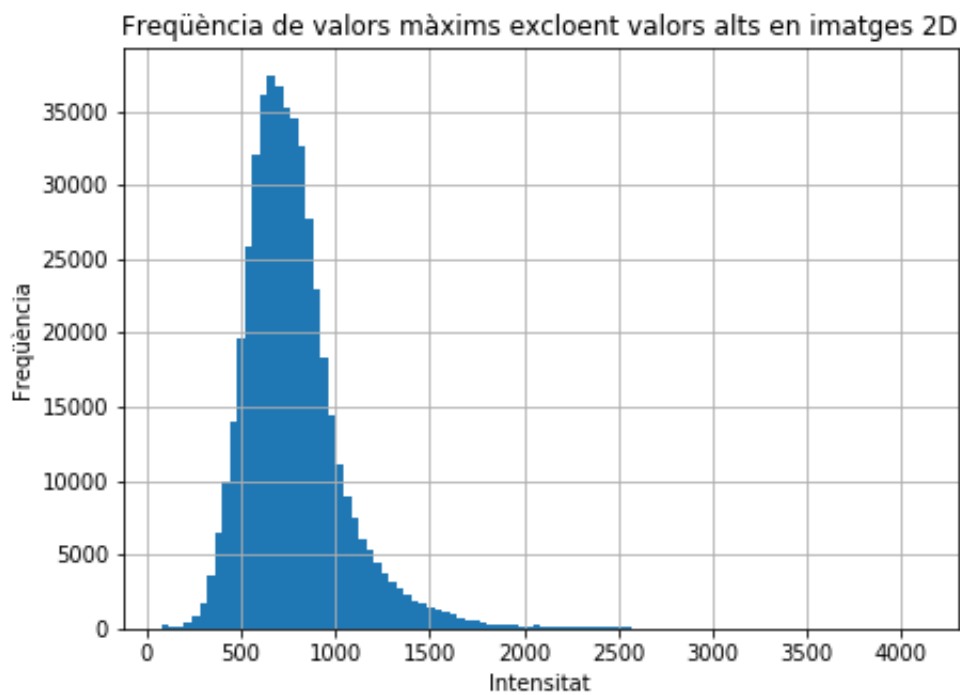


Figura A.2: Freqüència dels valors màxims d'intensitat de cadascuna de les imatges 2D del conjunt ADNI excloent aquells valors massa alts i poc freqüents.