



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIEROS
INDUSTRIALES VALENCIA

TRABAJO FIN DE GRADO EN INGENIERÍA BIOMÉDICA

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

AUTORA: CAROLINA GIMÉNEZ ARIAS

TUTOR: DAVID MORATAL PÉREZ

COTUTOR: RAFAEL ORTIZ RAMÓN

Curso Académico: 2017-18

AGRADECIMIENTOS

A mi familia, por estar siempre presentes a pesar de la distancia.

A Darwin, por su apoyo incondicional en todos mis proyectos.

A mis abuelos, por ser mi refugio durante estos años en Valencia.

A mi tutor David Moratal y a mi cotutor, Rafael Ortiz, por guiarme durante la elaboración de este trabajo.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

RESUMEN

La enfermedad de Alzheimer (EA) es una enfermedad neurodegenerativa que se caracteriza por un deterioro cognitivo y conductual de inicio insidioso y progresivo de aparición en la edad adulta, y que es responsable del 70 al 76% de los casos de demencia en los países desarrollados. Actualmente su diagnóstico se realiza en las fases más avanzadas de la enfermedad, por lo que su detección temprana es uno de los principales desafíos en la medicina actual. Existen muchos estudios que definen cuáles son los signos neuropatológicos clásicos de la enfermedad del Alzheimer, sin embargo, aún no existe un tratamiento eficaz una vez se ha desarrollado la enfermedad. De esta idea surge la importancia de encontrar biomarcadores en la fase prodrómica (fase sintomática) o incluso en la fase preclínica (fase asintomática) de la enfermedad que permitan identificar un futuro desarrollo de la EA, con el fin de aplicar terapias que puedan variar la evolución natural de la EA o incluso prevenir su aparición. Hasta el momento, el conocimiento de los principales hallazgos neuropatológicos y fisiopatológicos de la EA ha permitido detectar algunas de las proteínas implicadas en la aparición de la enfermedad (beta-amiloide 42, tau total y tau fosforilada). Asimismo, se sabe que las estructuras del lóbulo temporal medial, concretamente el hipocampo y el córtex entorrinal, se afectan precozmente en la EA. Estudios con resonancia magnética (RM) cuantitativa han sugerido que la atrofia hipocámpica se inicia incluso con anterioridad al inicio de la demencia y aumenta con la progresión de la enfermedad, por lo que podría convertirse en un buen biomarcador.

En los últimos años, el análisis cuantitativo y cualitativo de las imágenes de RM con el objetivo de extraer y evaluar la mayor cantidad de datos o parámetros posible para establecer biomarcadores de la enfermedad o lesión bajo estudio se ha convertido una de las líneas de investigación más prometedoras. El objetivo de este análisis reside en facilitar el diagnóstico médico, la predicción de la respuesta a tratamientos o la monitorización del estado de enfermedades. Una de las técnicas de extracción y análisis de parámetros más usadas en los últimos años ha sido el análisis de texturas, mediante el cual se evalúan matemáticamente las imágenes médicas a partir de la distribución de los píxeles o niveles de gris para hallar características no perceptibles a simple vista. El análisis de texturas se ha empleado sobretodo en estudios oncológicos (caracterización de los diferentes tipos de cáncer o identificación de los tipos de tumores cerebrales, por ejemplo), pero también se ha aplicado en otras áreas de investigación médica como las enfermedades cardiovasculares o las enfermedades neurodegenerativas, aunque en menor medida.

Además de la extracción de los parámetros que describen la imagen, el análisis de texturas implica la evaluación inteligente de estos parámetros. Para ello se emplean técnicas de aprendizaje máquina o *machine learning*. Estas combinan las disciplinas de la computación y la estadística con el fin de encontrar en un conjunto de datos correlaciones o patrones determinantes que permitan caracterizar o clasificar grupos de muestras (clases) o pronosticar comportamientos futuros en un sistema. Para ejecutar correctamente las técnicas de aprendizaje máquina hay que tener en cuenta varios aspectos como, por ejemplo, cuántas clases tenemos, qué modelo clasificador se adapta mejor a los datos, cuántos parámetros son necesarios para conseguir el mejor resultado o si es necesario aplicar métodos de validación cruzada para evaluar los resultados correctamente.

En el presente Trabajo de Fin de Grado se pretende analizar un conjunto parámetros de textura extraídos de imágenes de RM cerebrales empleando técnicas de aprendizaje máquina con el fin de

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

generar y estudiar modelos basados en estas texturas que permitan identificar las diferentes fases de la EA. Para ello, nos centraremos en análisis de texturas en 2D y 3D del hipocampo de tres grupos de sujetos: pacientes control (fase asintomática), pacientes con deterioro cognitivo leve (fase sintomática) y pacientes con EA en fase avanzada. El objetivo final del proyecto consistirá en crear un software intuitivo en el que poder analizar los diversos factores a tener en cuenta a la hora de aplicar técnicas de aprendizaje máquina (evaluación de distintos modelos, selección de características, validación cruzada, etc.) con el fin de obtener modelos óptimos de clasificación para la EA.

Palabras clave: enfermedad de Alzheimer (EA), análisis de texturas, aprendizaje máquina, imagen de resonancia magnética.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

ABSTRACT

Alzheimer's disease (AD) is a neurodegenerative disease characterized by a cognitive and behavioral impairment with an insidious and progressive onset during adulthood, and it is responsible for 70 to 76% of cases of dementia in developed countries. Currently, its diagnosis is made in the most advanced stages of the disease, so its early diagnosis is one of the main challenges in current medicine. There are many studies that define which are the classic neuropathological signs of Alzheimer's disease but, however, there is no effective treatment once the disease has developed. This is the main reason for exploring biomarkers in the prodromal phase (symptomatic phase) or even in the preclinical phase (asymptomatic phase) of the disease ; it would allow the identification of a future development of AD in order to apply therapies that could change the natural evolution of the disease or even prevent its appearance. So far, the knowledge of neuropathological and physiopathological findings has allowed detecting some of the proteins involved in the appearance of the disease (amyloid beta 42, total tau and phosphorylated tau). It is also known that the structures of the medial temporal lobe, specifically the hippocampus and the entorhinal cortex, are also affected in early AD. Some quantitative studies with magnetic resonance imaging (MRI) have suggested that the hippocampal atrophy begins even before the onset of dementia and increases with the progression of the disease, so it could become a good biomarker.

In recent years, the quantitative and qualitative analysis of MRI with the aim of extracting and evaluating information from these images to establish biomarkers of the disease or lesion under analysis has become one of the most important lines of research. The objective of this analysis is to help in the medical diagnosis, the prediction of the response to treatments or the monitoring of the stage of diseases. In the last years, the analysis of textures has become one of the most common feature extraction and analysis techniques, through which medical images are mathematically evaluated using the pixels or gray levels distribution to find features that are not visually detectable. The analysis of textures has been used mainly in cancer studies (characterization of different types of cancer or identification of brain tumours types, for example), but it is also applied to a lesser extent in other areas of medical research as cardiovascular or neurodegenerative diseases.

In addition to the extraction of the parameters that describe the image, texture analysis involves its intelligent evaluation. This evaluation is made with machine learning techniques. These techniques combine computational and statistical disciplines in order to find correlations in a data set which allow characterizing or classifying groups of samples (classes) and predict future behaviors of a system. To correctly execute these techniques, several aspects have to be taken into account, for example the number of classes that we have, which is the best classification model for our data set, the number of parameters which are needed to achieve the best result or if it is necessary to apply cross validation methods to evaluate the results.

The main objective of the present project is to analyze a set of texture parameters extracted from brain MRI using machine learning techniques in order to generate and study models based on these textures that will help in the identification of the different stages of AD. To that end, we will focus on 2D and 3D texture analysis of the hippocampus of three groups of subjects: control patients (asymptomatic phase), patients with mild cognitive impairment (symptomatic phase) and patients with advanced AD. To perform this analysis we will create an intuitive software which enables to analyze the factors to take into account when applying machine learning techniques (evaluation of

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

different models, feature selection, cross-validation, etc.) for the purpose of obtaining the optimal classification models for AD.

Keywords: Alzheimer disease (AD), texture analysis, machine learning, magnetic resonance image (MRI).

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

RESUM

La malaltia d'Alzheimer és una malaltia neurodegenerativa que es caracteritza per un deteriorament cognitiu i conductual d'inici insidiós i progressiu d'aparició en l'edat adulta, i que és responsable del 70 al 76% dels casos de demència en els països desenvolupats. Actualment el seu diagnòstic es realitza en les fases més avançades de la malaltia. Pel que la seua detecció precoç, és un dels principals reptes en la medicina actual. Hi ha molts estudis que defineixen quins són els signes neuropatològics clàssics de la malaltia de l'Alzheimer, però, encara no existeix un tractament eficaç una vegada s'ha desenvolupat la malaltia. D'aquesta idea sorgeix la importància de trobar biomarcadors en la fase prodròmica (fase simptomàtica) o fins i tot en la fase preclínica (fase asimptomàtica) de la malaltia que permeten identificar un futur desenvolupament de la malaltia d'Alzheimer, per tal d'aplicar teràpies que puguin variar la evolució natural de la malaltia d'Alzheimer o fins i tot prevenir la seua aparició. Fins al moment, el coneixement dels principals descobriments neuropatològics i fisiopatològics de la malaltia d'Alzheimer ha permès detectar algunes de les proteïnes implicades en l'aparició de la malaltia (beta-amiloide 42, tau total i tau fosforilada). Així mateix, se sap que les estructures del lòbul temporal medial, concretament l'hipocamp i el còrtex entorrinal, s'afecten precoçment en la malaltia d'Alzheimer. Estudis amb ressonància magnètica quantitativa han suggerit que l'atròfia hipocàmpica s'inicia fins i tot abans de l'inici de la demència i augmenta amb la progressió de la malaltia, de manera que podria esdevenir un bon biomarcador.

En els últims anys, l'anàlisi quantitativa i qualitativa de les imatges de ressonància magnètica amb l'objectiu d'extreure i avaluar la major quantitat de dades o paràmetres possible per establir biomarcadors de la malaltia o lesió sota estudi s'ha convertit una de les línies d'investigació més prometedores. L'objectiu d'aquesta anàlisi consisteix en facilitar el diagnòstic mèdic, la predicció de la resposta a tractaments o la monitorització de l'estat de malalties. Una de les tècniques d'extracció i anàlisi de paràmetres més usades en els últims anys ha estat l'anàlisi de textures, mitjançant la qual s'avaluen matemàticament les imatges mèdiques a partir de la distribució dels píxels o nivells de gris per trobar característiques no perceptibles a simple vista. L'anàlisi de textures s'ha emprat sobretot en estudis oncològics (caracterització dels diferents tipus de càncer o identificació dels tipus de tumors cerebrals, per exemple), però també s'ha aplicat en altres àrees d'investigació mèdica com les malalties cardiovasculars o les malalties neurodegeneratives, encara que en menor grau.

A més de l'extracció dels paràmetres que descriuen la imatge, l'anàlisi de textures implica l'avaluació intel·ligent d'aquests paràmetres. Per a això s'empren tècniques d'aprenentatge màquina o *machine learning*. Aquestes combinen les disciplines de la computació i l'estadística per tal de trobar en un conjunt de dades correlacions o patrons determinants que permetin caracteritzar o classificar grups de mostres (*classes*) o pronosticar comportaments futurs en un sistema. Per executar correctament les tècniques d'aprenentatge màquina cal tenir en compte diversos aspectes com, per exemple, quantes classes tenim, quin model classificador s'adapta millor a les dades, quants paràmetres més rellevants per aconseguir el millor resultat o si cal aplicar mètodes de validació creuada per avaluar els resultats correctament.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

En el present Treball de Fi de Grau es pretén analitzar un conjunt de paràmetres de textura extrets d'imatges de RM cerebrals emprant tècniques d'aprenentatge màquina per tal de generar i estudiar models basats en aquestes textures que permeten identificar les diferents fases de la malaltia d'Alzheimer. Per a això, ens centrarem en l'anàlisi de textures en 2D i 3D de l'hipocamp de tres grups de subjectes: pacients control (fase asimptomàtica), pacients amb deteriorament cognitiu lleu (fase simptomàtica) i pacients amb malaltia d'Alzheimer en fase avançada. L'objectiu final del projecte consistirà en crear un programari intuïtiu en el qual poder analitzar els diversos factors a tenir en compte a l'hora d'aplicar tècniques d'aprenentatge màquina (avaluació de diferents models, selecció de característiques, validació creuada, etc.) amb el fi d'obtenir models òptims de classificació per a la malaltia d'Alzheimer.

Paraules clau: malaltia d'Alzheimer, anàlisi de textures, aprenentatge automàtic, imatge de ressonància magnètica.

DOCUMENTOS CONTENIDOS EN EL TFG

- Documento memoria
- Documento presupuesto

DOCUMENTO

MEMORIA

**ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE
TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA
DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE
ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA**

AUTORA: CAROLINA GIMÉNEZ ARIAS

TUTOR: DAVID MORATAL PÉREZ

Centro de Biomateriales e Ingeniería Tisular, Universitat Politècnica de València

COTUTOR: RAFAEL ORTIZ RAMÓN

Centro de Biomateriales e Ingeniería Tisular, Universitat Politècnica de València

CURSO 2017-18

ÍNDICE MEMORIA

1. MOTIVACIÓN Y OBJETIVOS.....	5
2. INTRODUCCIÓN	6
2.1. Demencia.....	6
2.1.1 Definición y prevalencia	6
2.1.2 Fases	6
2.1.3 Diagnóstico	8
2.1.4 Deterioro cognitivo leve.....	10
2.1.5 Alzheimer.....	11
2.2. Hipocampo.....	12
2.3. Neuroimagen	14
2.3.1 Imagen por Resonancia Magnética.....	14
2.3.2 Biomarcadores	15
2.3.3 Análisis de texturas.....	16
2.4 Métodos de aprendizaje máquina	18
2.4.1 Métodos de clasificación.....	19
2.4.1.1 <i>K-Nearest Neighbors</i>	19
2.4.1.2 <i>Support Vector Machine</i>	21
2.4.1.3 <i>Decision Trees</i>	23
2.4.2 Métodos de selección de características	24
2.4.2.1 <i>P</i> -valor.....	25
2.4.2.2 Fisher	26
2.4.2.3 Relief y Relief-F	26
2.4.3 Validación de sistemas	27
2.4.3.1 Validación <i>hold-out</i>	27
2.4.3.2 Validación cruzada	28
2.4.4 Métodos de evaluación de resultados.....	29
2.4.4.1 Matriz de confusión	30
2.4.4.2 Curva ROC y AUC.....	31
3. MATERIALES.....	33
3.1 Datos.....	33
3.2 Software ALTEA	34
4. MÉTODOS	35
4.1 Segmentación y análisis de texturas	35
4.2 Visualización de datos	36
4.3 Clasificación	39
5. RESULTADOS Y DISCUSIÓN	46
5.1 Desarrollo de herramienta para la visualización de datos	46
5.1.1 Visualización texturas 2D	46
5.1.2. Visualización texturas 3D	48
5.2 Desarrollo <i>Classifiers</i>	49
5.3 Análisis de texturas.....	49
5.3.1. Análisis de texturas 2D	50
5.3.2. Análisis de texturas 3D	55
5.3.3. Comparación análisis 2D y 3D.....	60
5.4 Estudio de los métodos	61
5.5 Discusión con otros estudios	63
6. CONCLUSIONES.....	64
7. LÍNEAS FUTURAS.....	65
8. REFERENCIAS	66

ÍNDICE FIGURAS

Figura 1. De izquierda a derecha: localización del hipocampo y la corteza entorrinal en el cerebro, corte coronal del lóbulo temporal medial donde se observa la forma del hipocampo y figura de un caballito de mar, de donde obtiene su nombre debido a su similitud	12
Figura 2. De izquierda a derecha: diferencia hipocampo entre paciente control, deterioro cognitivo leve y paciente con enfermedad de Alzheimer.....	13
Figura 3. Esquema del procedimiento de obtención de una IMR	15
Figura 4. Tipos de función discriminante.	18
Figura 5. Ilustración de las fronteras y regiones definidas por modelos K-vecinos con distintos valores de K.	20
Figura 6. Clasificación con SVM.	21
Figura 7. Ejemplo de aplicación de árbol de decisión top-down para clasificar frutas (hojas) en función de características como el tamaño o el color (nodos).....	23
Figura 8. Método de validación <i>hold-out</i>	27
Figura 9. Método de validación cruzada con K iteraciones (en este caso K=4).	28
Figura 10. Ejemplo de curvas ROC	31
Figura 11. Pantalla inicial gráfica de dispersión.	37
Figura 12. Carga de datos en la herramienta.	38
Figura 13. Visualización de ejemplo de gráfica de dispersión en función de la varianza y la curtosis para los datos de la tabla ROI_1_D_2D.....	38
Figura 14. Pantalla inicial del clasificador.....	39
Figura 15. Pantalla de opción de clasificador.....	40
Figura 16. Pantalla de la opción KNN que solicita el número de vecinos.	40
Figura 17. Pantalla de la opción KNN que solicita la métrica para calcular la distancia.	40
Figura 18. Pantalla de la opción SVM que solicita la función kernel deseada para el modelo.	41
Figura 19. Pantalla de la opción SVM que solicita el número de restricciones.	41
Figura 20. Pantalla de la opción DT que solicita el número máximo de divisiones.....	42
Figura 21. Opción de selección de las clases del problema.....	42
Figura 22. Opción de selección del método de ponderación de características	43
Figura 23. Opción de selección del método de validación.	43
Figura 24. Clasificador en funcionamiento.....	44
Figura 25. Ejemplo matriz de confusión.....	45
Figura 26. Diagrama de dispersión de <i>Variance_global</i> y <i>Kurtosis_Global</i> para ROI_2_I_2D.....	47
Figura 27. Diagrama de dispersión de <i>Variance_global</i> y <i>Kurtosis_Global</i> para ROI_1_I_2D.....	47
Figura 28. Diagrama de dispersión de <i>Variance_global</i> y <i>Kurtosis_Global</i> para ROI_2_I_3D.....	48
Figura 29. Diagrama de dispersión de <i>Variance_global</i> y <i>Kurtosis_Global</i> para ROI_1_D_3D.....	48
Figura 30. Visualización resultados de regiones de interés de 2 dimensiones.....	52
Figura 31. Matriz de confusión para la clasificación EMCI vs. AD de la ROI_2_I_2D.	53
Figura 32. Curva <i>profile</i> y curva ROC para la clasificación EMCI vs. AD de la ROI_2_I_2D.....	53
Figura 33. Matriz de confusión para la clasificación multiclase de ROI_2_I_2D.....	54
Figura 34. Visualización resultados de regiones de interés de 2 dimensiones.....	57
Figura 35. Matriz de confusión para la clasificación multiclase de ROI_1_D_3D.	58
Figura 36. Matriz de confusión para la clasificación EMCI vs. AD de la ROI_3_I_3D.	59

**ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA
DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA**

Figura 37. Curva <i>profile</i> y curva ROC para la clasificación EMCI vs. AD de la ROI_3_I_3D.....	59
Figura 38. Comparación de los resultados 2D y 3D para una ROI de 8 píxeles.....	60
Figura 39. Frecuencia de uso de métodos de clasificación para resultados 2D.....	61
Figura 40. Frecuencia de uso de métodos de clasificación para resultados 3D.....	62
Figura 41. Frecuencia de uso de métodos de ranking para resultados 3D.	62

ÍNDICE TABLAS

Tabla 1. Evolución de la severidad de la demencia en escala GDS según años	8
Tabla 2. Estructura de matriz de confusión para un problema de clasificación binario.....	30
Tabla 3. Texturas obtenidas para cada ROI.....	36
Tabla 4. Parámetros para elaborar los modelos analizados	49
Tabla 5. Resultados para regiones de interés de 2 dimensiones.....	49
Tabla 6. Resultados para regiones de interés de 3 dimensiones.....	54

1. MOTIVACIÓN Y OBJETIVOS

El desarrollo tecnológico y los avances en investigación de los últimos años han permitido la implementación de numerosas mejoras dentro del ámbito sanitario, incluyendo los procesos de detección y terapia de múltiples enfermedades que afectan a la población a nivel mundial. Sin embargo, dentro de este prometedor contexto, aún existen enfermedades cuyo diagnóstico sigue siendo fundamentalmente clínico, es decir, que no son diagnosticadas hasta que el propio paciente padece sus síntomas, como es el caso de la enfermedad del Alzheimer.

La enfermedad de Alzheimer se caracteriza por ser un tipo de demencia de inicio gradual que aparece en edades adultas, es por ello que, como consecuencia del aumento del envejecimiento poblacional, son cada vez más las personas afectadas por esta patología. Existen actualmente numerosos estudios que se centran en tratar de encontrar un método que permita la detección de esta enfermedad en sus fases más tempranas, de modo que se comience con una terapia acorde a las características de la enfermedad también de forma temprana.

Para tratar de detectar precozmente esta patología resulta crucial el estudio de imágenes de resonancia magnética (IRM) cerebral, pues se tienen indicios de que la región del hipocampo es una de las primeras estructuras cerebrales en verse afectadas por la atrofia neuronal propia del Alzheimer (Wortmann, 2012). Las técnicas de análisis de texturas permiten extraer numerosas características de las IRM que resultan de elevada utilidad a la hora de encontrar un biomarcador que facilite la detección y el seguimiento del progreso de la patología del Alzheimer en concreto (Freeborough y Fox, 1998). Además, existe un interés creciente en combinar esta metodología de estudio del Alzheimer con métodos de aprendizaje máquina, de modo que, mediante estas técnicas de inteligencia artificial se entrene un sistema capaz de discriminar las imágenes sanas de las patológicas y diagnosticar así la enfermedad de Alzheimer (Zhang *et al.*, 2012). A su vez, muchos de estos estudios tratan de evaluar la influencia de factores como la dimensionalidad sobre los resultados de la clasificación (Martí-Climent *et al.*, 2010).

Partiendo de esta premisa, a lo largo de este trabajo y como principal objetivo, se pretenden analizar texturas extraídas a partir de IRM empleando técnicas de aprendizaje máquina con el fin de encontrar marcadores significativos que permitan la detección de Alzheimer en fases tempranas. Para llevar a cabo esta tarea, en primer lugar, se segmentará el hipocampo de tres grupos diferenciados de pacientes a partir de imágenes de resonancia magnética potenciadas en T1: controles sanos, pacientes con enfermedad de Alzheimer, y pacientes con deterioro cognitivo leve. A partir de dicha segmentación, se extraerán distintas texturas del hipocampo. Se emplearán métodos de análisis de texturas con el propósito de obtener parámetros cuantitativos, tanto bidimensionales como tridimensionales que, posteriormente, serán particionados en dos subconjuntos de datos distintos: entrenamiento y prueba. Empleando dichos subconjuntos de datos, se entrenarán y testearán distintos modelos de evaluación de características para estudiar la capacidad predictiva de distintas texturas ante posibles casos de Alzheimer. A su vez, se tratará de evaluar cómo se ve afectada la clasificación en función del tamaño de la región de interés seleccionada para extraer las texturas del hipocampo y de si estas han sido extraídas bidimensional o tridimensionalmente.

2. INTRODUCCIÓN

2.1 Demencia

2.1.1 Definición y prevalencia

La demencia es un término general que permite describir un conjunto de síntomas neurodegenerativos que afectan principalmente a la memoria, la capacidad de razonamiento, la ejecución de actos complejos o las capacidades sociales, entre otros. El término *demencia* no constituye una enfermedad particular, sino que hace referencia al conjunto de síntomas consecuentes del deterioro cognitivo y conductual que interfieren en la vida diaria del enfermo, principalmente debido a que reducen significativamente su autonomía funcional.

La enfermedad del Alzheimer (EA) es el tipo de demencia más común. En la actualidad, es considerada como la causa más frecuente de demencia neurodegenerativa en los países desarrollados, cuya prevalencia se incrementa con la edad, esto hace que, aunque su etiología sea desconocida, la edad sea considerada como el principal factor de riesgo no modificable (Valls-Pedret, Molinuevo y Rami, 2010).

La importancia tanto a nivel sanitario como social de la demencia en los países desarrollados es tal que muchos autores la consideran como epidemia (Molinuevo Guix, 2018). Esto es debido a que, desde el pasado siglo, la población mundial ha experimentado un envejecimiento que se prevé que continúe aumentando. De este modo, se calcula que para el año 2045 el número de personas mayores superará la cantidad de niños y, como consecuencia, la demencia cobrará aún más importancia en nuestra sociedad (Sosa-Ortiz, Acosta-Castillo y Prince, 2012). En la actualidad, se estima que la prevalencia de esta patología en la población de 65 a 70 años en España y Europa oscila entre el 9,1 y el 13,9%. Además, se calcula que entre el 0,6 y el 1% de las personas entre 65 y 69 años de edad y entre el 20 y 30% de las personas con edades iguales o superiores a 90 años se ven afectadas por el síndrome demencial (Molinuevo Guix, 2018).

2.1.2 Fases

La progresión de la demencia en los individuos que la padecen debe ser estudiada para localizar en qué estadio de deterioro cognitivo se encuentra el caso en concreto. Para ello, se estudian los síntomas del paciente, de modo que, de acuerdo al grado de desarrollo, se facilite el tratamiento determinado.

La demencia como tal no se desarrolla del mismo modo en todos los pacientes, sino que su evolución es variable, pues puede verse influenciada por distintos factores, es por ello que, definir la fase de la demencia de un individuo puede llegar a ser una tarea compleja. En ocasiones, se trata de simplificar esta tarea diferenciando tres únicos estados de desarrollo o progresión de la enfermedad: “etapa temprana”, “etapa media” y “etapa avanzada”, sin embargo, en otros casos es necesario describir una fase más exacta que defina los síntomas del paciente, para ello existen varias escalas entre las cuales destaca la Escala Global del Deterioro (*Global Deterioration Scale*, GDS) también conocida como la Escala de Reisberg.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

La Escala Global del Deterioro para la evaluación de la demencia primaria degenerativa divide el progreso de la enfermedad en siete etapas en función de la declinación cognitiva del paciente (Reisberg *et al.*, 1982)(Reisberg *et al.*, 1982).

- Fase 1: Ningún deterioro cognitivo.
- Fase 2: Deterioro cognitivo muy leve.
- Fase 3: Deterioro cognitivo leve.
- Fase 4: Deterioro cognitivo moderado.
- Fase 5: Deterioro cognitivo moderadamente severo.
- Fase 6: Deterioro cognitivo severo.
- Fase 7: Deterioro cognitivo muy severo.

Las tres primeras etapas de esta escala no son diagnosticadas como demencia debido a que durante la fase 1 la persona no experimenta pérdidas de memoria, y durante las fases 2 y 3 sí que existe una pérdida de memoria, aunque es de carácter leve. En concreto, la fase 2 se corresponde con el deterioro cognitivo normal asociado a la edad, por el cual se ve afectada más de la mitad de la población mayor a 65 años. Mientras tanto, la fase 3 o deterioro cognitivo leve, también conocido por MCI por sus siglas en inglés (*Mild Cognitive Impairment*) suele conllevar una pérdida creciente de memoria a la vez que se comienzan a manifestar síntomas de dificultad de concentración o ansiedad. Esta etapa es de difícil detección, sobre todo en personas que no realizan tareas ocupacionales o sociales complejas, pues los síntomas pueden no ser evidentes para los familiares y personas del entorno del paciente. Esta fase tiene una duración media de 7 años antes del inicio de la demencia, aunque en la mayoría de las personas se comienzan a manifestar síntomas claros de demencia en un intervalo de 2 a 4 años.

En cuanto a la fase 4, esta sería considerada como demencia en una “etapa temprana”, pues incluye una disminución evidente de la habilidad de concentración del paciente. Las personas que se encuentran en este estadio pueden sobrevivir de forma independiente, aunque requieren de ayuda externa para la realización de actividades complejas de la vida cotidiana. Las fases 5 y 6 se corresponderían con una “etapa media” de demencia, durante la fase 5 el paciente comienza a requerir ayuda para completar sus actividades habituales, mientras que durante la sexta fase el paciente requiere de ayuda extensiva para realizar dichas actividades y no tiene recuerdos de los eventos recientes, además es común la aparición de cambios en la personalidad y ansiedad.

Por último, la fase 7 se correspondería con una demencia severa, los pacientes cuentan con una total dependencia a la hora de realizar cualquier actividad, además de que pierden la capacidad de comunicarse. La duración promedio de esta etapa es de 2,5 años.

Las duraciones de cada una de las etapas diferenciadas por el grado de severidad quedan recogidas en la Tabla 1. Cabe destacar que la duración promedio de la vida de un paciente con enfermedad de Alzheimer una vez aparecen sus síntomas es de 10 años, aunque dicha expectativa de vida se ve disminuida en las formas de inicio precoz, es decir, si la enfermedad se desarrolla en el paciente antes de los 60 años.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Diagnóstico	Fase	Duración promedia
Sin demencia	Fase 1: Normal	
	Fase 2: Deterioro cognitivo muy leve	
	Fase 3: Deterioro cognitivo leve	7 años
Demencia en etapa temprana	Fase 4: Deterioro cognitivo moderado	2 años
Demencia en etapa media	Fase 5: Deterioro cognitivo moderadamente severo	1,5 años
	Fase 6: Deterioro cognitivo severo	2,5 años
Demencia en etapa avanzada	Fase 7: Deterioro cognitivo muy severo	2,5 años

Tabla 1. Evolución de la severidad de la demencia en escala GDS según años. *Fuente: Fisher Center for Alzheimer's Research Foundation. (Clinical Stages of Alzheimer's | Fisher Center for Alzheimer's Research Foundation, fecha último acceso: 09-05-2018).*

2.1.3 Diagnóstico

A lo largo de los últimos años, se han tratado de estudiar y desarrollar criterios cada vez más refinados con el fin de promover una identificación precoz de la demencia, y en concreto, de la enfermedad de Alzheimer. Este interés se debe principalmente a que se conoce que los tratamientos que modifican el curso de estas enfermedades neurodegenerativas son más eficaces si son aplicados en etapas tempranas.

El diagnóstico clínico de la demencia se basa en definiciones que se encuentran principalmente contenidas en el Grupo de Trabajo de la Asociación Nacional de Neurología, Trastornos Comunicativos y Accidentes Cerebrovasculares y Trastornos Relacionados (NINCDS-ADRDA) y la 3ª y 4ª edición del Manual de Diagnóstico y Estadística (DSM-IIR y DSM-IV) (Knopman *et al.*, 2001).

Las labores de detección y clasificación de cualquier tipo de demencia son complejas, pues actualmente no existe ninguna prueba definitiva que confirme la presencia de esta enfermedad. Para poder llevar a cabo este diagnóstico, es necesario que al menos dos funciones cerebrales básicas se encuentren alteradas de modo que obstaculicen la vida cotidiana del paciente. Estas funciones serían: la memoria, la facultad del lenguaje, la percepción visual y las capacidades para concentrarse, razonar y resolver problemas.

El procedimiento común que se sigue a la hora de realizar el diagnóstico de cualquier deterioro cognitivo, antes de aplicar los criterios diagnósticos de tipos específicos de demencia se basa en que, en primer lugar, se examina al paciente para determinar si existe la presencia de algún tipo de deterioro cognitivo. Para ello, se lleva a cabo la anamnesis, es decir, se recogen una serie de datos en la historia clínica de este donde figure su información básica, la presencia de factores de riesgo vasculares, los antecedentes familiares y el nivel educativo del paciente. Además, se llevará a cabo un examen clínico del paciente para evaluar su estado físico. A su vez, se realizará una evaluación neurológica que trate de detectar la presencia de afasia (trastorno del lenguaje que dificulta la comunicación hablada), apraxia (trastorno que incapacita al paciente a la hora de llevar

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

a cabo movimientos cuando se le solicita), agnosia (trastorno que incapacita la identificación de estímulos previamente aprendidos) o trastornos de la función ejecutiva como alteraciones del estado de ánimo o delirios. Una vez se tienen sospechas, se trata de descartar otras patologías con sintomatología similar, como puede ser el deterioro cognitivo leve.

A su vez, se realizan otra serie de pruebas a distintos niveles para confirmar la presencia de demencia:

- Pruebas cognitivas y neuropsicológicas. Estas pruebas miden las funciones cognitivas y las capacidades mentales del paciente como, por ejemplo, la memoria, el lenguaje o el razonamiento.
- Evaluación psiquiátrica. Se evalúan si síntomas como la depresión u otras afecciones psiquiátricas contribuyen a los síntomas observados e interfieren con el diagnóstico.
- Análisis de laboratorios. Estas pruebas tratan de determinar la presencia en sangre u orina de algunas sustancias como la glucosa, la vitamina B12 o algunas hormonas tiroideas, puesto que alteraciones significativas en la presencia de dichas sustancias en el organismo pueden ser síntoma de desorden neurológico (Knopman *et al.*, 2001).
- Pruebas de neuroimagen y trazados gráficos. Las técnicas de neuroimagen cobran una elevada importancia en el estudio de deterioros cognitivos ya que permiten el estudio cerebral no sólo a nivel estructural, sino a nivel funcional, lo que facilita la comprensión del comportamiento de la demencia como enfermedad. Las pruebas más usadas son:
 - Tomografía Axial Computerizada (TAC) o Resonancia Magnética Nuclear (RMN). Mediante estos estudios estructurales se buscan evidencias de accidentes cerebrovasculares, tumores e hidrocefalias, que pueden llegar a ser precursores de determinados tipos de demencia.
 - Tomografía por Emisión de Positrones (PET) o Tomografía Computerizada de Emisión de Fotón (SPECT). Se trata de pruebas que evalúan la funcionalidad del cerebro y que permiten observar patrones en su actividad o incluso la deposición de proteínas como la β -amiloide, relacionada con el Alzheimer (Jiménez Bonilla y Carril Carril, 2013).

Tras confirmar la presencia de demencia, se evalúa el grado de deterioro cognitivo, para ello, se emplean escalas como la anteriormente mencionada GDS o la escala FAST, acrónimo del término inglés *Functional Assessment Staging*, mediante la cual se determina la fase de demencia evaluando el grado de funcionalidad del paciente a la hora de realizar sus actividades cotidianas.

Como ya se ha comentado, el principal inconveniente reside en que, para poder diagnosticarla, la demencia debe encontrarse en fases avanzadas, puesto que no es posible predecir la evolución de un caso de deterioro cognitivo leve, no diagnosticado como demencia, al deterioro cognitivo moderado, considerado como demencia, y por lo tanto los tratamientos para modificar el curso de la enfermedad serán menos efectivos. Cabe destacar que, el diagnóstico temprano de cualquier demencia a su vez resulta de elevada importancia a la hora de planificar una terapia que facilite el padecimiento de esta enfermedad, de modo que se fomente principalmente la comodidad del propio paciente y a su vez la de sus familiares, pues entre otras cosas permite que este tenga potestad en la toma de decisiones sobre sus cuidados futuros.

2.1.4 Deterioro cognitivo leve

El deterioro cognitivo leve (DCL) o *mild cognitive impairment* (MCI) en inglés, representa un estadio intermedio entre el deterioro cognitivo normal consecuente del envejecimiento y la demencia, que afecta principalmente al área de la memoria. Se trata de una afección que ha sido ampliamente investigada, considerada como trastorno cognitivo, a pesar de que no influya sobre las actividades cotidianas del enfermo y no cumpla con los criterios para ser considerada como demencia (Sosa-Ortiz, Acosta-Castillo y Prince, 2012). Según la escala GDS este tipo de deterioro se corresponde con la fase 3.

Los criterios originales de diagnóstico del deterioro cognitivo leve apuntaban la importancia del deterioro de la memoria preservando otros dominios cognitivos como la realización de actividades de la vida cotidiana del paciente. Estos criterios orientaban el diagnóstico hacia una etapa prodrómica de la enfermedad de Alzheimer, es decir, hacia una etapa inicial de dicha enfermedad. Sin embargo, posteriormente se determinó que no todas las formas de DCL evolucionaban a la enfermedad de Alzheimer, por lo que esta forma de diagnóstico requería de una conceptualización más amplia. En 2003 se propusieron una serie de nuevos criterios más específicos que permiten describir cuatro tipos de DCL (Pose y Manes, 2010):

- DCL amnésico dominio único
- DCL amnésico dominio múltiple
- DCL no amnésico dominio único
- DCL no amnésico dominio múltiple

En cuanto a la prevalencia de DCL, existen estudios que confirman que, en la población de individuos mayores de 70 años la tasa de prevalencia de este tipo de deterioro cognitivo oscila entre el 14 y el 18% (Pose y Manes, 2010). Las cifras de incidencia varían anualmente del 8 a 77 por cada 1000 personas.

Además, el riesgo de que pacientes con DCL acaben desarrollando algún tipo de demencia también presenta una enorme variabilidad en función del estudio realizado y la fuente de pacientes. Sin embargo, a pesar de estas diferencias, es importante resaltar que la tasa de progresión a demencia de sujetos con DCL mayores de 65 años es mucho más elevada que la calculada para sujetos mayores de 65 sin ningún deterioro cognitivo. Cabe destacar que el DCL amnésico multidominio es el subtipo con mayor tasa de conversión a demencia (Ahmed *et al.*, 2008).

El diagnóstico del deterioro cognitivo leve, al igual que el diagnóstico demencial, no es un proceso sencillo. No basta con realizar una serie de pruebas cognitivas, sino que a su vez se requiere de estudios de la anamnesis del paciente, neuroimágenes, evaluaciones neuropsicológicas e incluso la opinión de familiares o personas que pertenezcan al entorno del paciente.

Una de las principales líneas de estudio del DCL se basa en encontrar métodos que logren predecir la progresión de deterioro cognitivo leve a demencia. Existen indicios que apuntan a que el mejor modelo predictivo es aquel que combina las neuroimágenes y otros biomarcadores con las diferentes fases del proceso patológico puesto que no existe ningún marcador que, individualmente, tenga capacidad predictiva uniformemente a lo largo del transcurso de la

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

enfermedad (Jack *et al.*, 2008). Gracias a las técnicas de neuroimagen estructurales, se conoce que aparece una reducción del volumen del lóbulo temporal medial, concretamente en el hipocampo en individuos con DLC, en comparación con sujetos de control sin ningún tipo de deterioro. A su vez, gracias a las técnicas de imagen funcional, también se han podido llevar a cabo otros estudios que revelaron una relación entre el hipometabolismo en regiones temporales y parietales en los pacientes con DCL y la progresión de estos a demencia. Además, mediante estas imágenes se pueden detectar in vivo depósitos amiloides, los cuales se cree que guardan relación con la progresión de los individuos con DCL a demencia.

2.1.5 Alzheimer

La enfermedad de Alzheimer (EA) es una enfermedad neurodegenerativa caracterizada por la presencia de deterioro intelectual y conductual de inicio repentino y curso progresivo que aparece durante la edad adulta, principalmente en la etapa de vejez (Valls-Pedret, Molinuevo y Rami, 2010). Fue descrita por primera vez en 1907 por el doctor Alois Alzheimer y hoy en día es considerada como la causa más frecuente de demencia neurodegenerativa en los países más desarrollados. Se calcula que la enfermedad de Alzheimer es responsable del 70 al 76% de los casos de demencia en los países desarrollados (Molinuevo Guix, 2018).

Su etiología es desconocida, aunque se piensa que probablemente tenga una causa multifactorial. Los principales cambios neuropatológicos de la enfermedad de Alzheimer se centran en la pérdida de sinapsis neuronal debida a la agregación anormal de las proteínas implicadas en la enfermedad del líquido cefalorraquídeo (LCR): tau hiperfosforilada intraneuronal en forma de ovillos neurofibrilares, tau total y β -amiloide extraneuronal en forma de placas seniles. Esta pérdida de sinapsis y muerte neuronal es consecuencia del desequilibrio crónico que se genera entre la producción y la eliminación de la proteína β -amiloide, como consecuencia, esta se acumula, desatando una cascada que finaliza con la muerte neuronal. Los pacientes con EA presentan una concentración disminuida de β -amiloide 42 y un incremento de los niveles de tau total y tau fosforilada en el LCR.

Actualmente, la detección de la EA se basa en la aplicación de criterios clínicos, ya que de momento no existen biomarcadores confiables para llevar a cabo esta tarea. Mediante estos criterios es posible diagnosticar la EA como probable, en caso de que exista presencia de demencia de inicio insidioso y progresivo en ausencia de otras enfermedades sistémicas o cerebrales que puedan explicar el trastorno cognitivo. Por otro lado, para poder confirmar un diagnóstico definitivo de Alzheimer, se precisa de un examen neuropatológico del paciente, tanto a nivel macroscópico, donde aparece una atrofia cerebral que predomina en la corteza y se hace más pronunciada en el córtex temporal medial, en concreto en el hipocampo y el córtex entorrinal, como a nivel microscópico, donde se puede observar la pérdida de neuronas, así como la aparición de placas amiloideas y ovillos neurofibrilares, característicos de la EA (Molinuevo Guix, 2018). Además, estudios demuestran que la patología neuronal comienza incluso hasta una década antes de que la sintomatología clínica empiece. Todo esto implica que, para poder diagnosticar la EA, debe existir un daño neuropatológico importante que afecte de forma generalizada a múltiples áreas cerebrales, incluyendo áreas cognitivas, lo que dificulta enormemente la modificación del curso evolutivo de esta enfermedad (Valls-Pedret, Molinuevo y Rami, 2010).

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

En cuanto a su tratamiento, no existe ningún recurso farmacológico o terapia capaz de detener la evolución de la enfermedad de Alzheimer, sin embargo, durante las fases temprana y media de la evolución de esta enfermedad se pueden suministrar medicamentos al paciente que prevengan el empeoramiento de algunos de los síntomas o conductas como la ansiedad o el insomnio, de modo que se garantice la comodidad para el propio paciente y se facilite su cuidado.

2.2 Hipocampo

El hipocampo se corresponde con una región del cerebro localizada en el interior de la parte medial o interna del lóbulo temporal, bajo la superficie cortical. El concepto de hipocampo fue propuesto por el anatomista Giulio Cesare Aranzi durante el siglo XVI, quien observó en un corte coronal del cerebro humano el parecido de esta estructura a un “caballito de mar”.

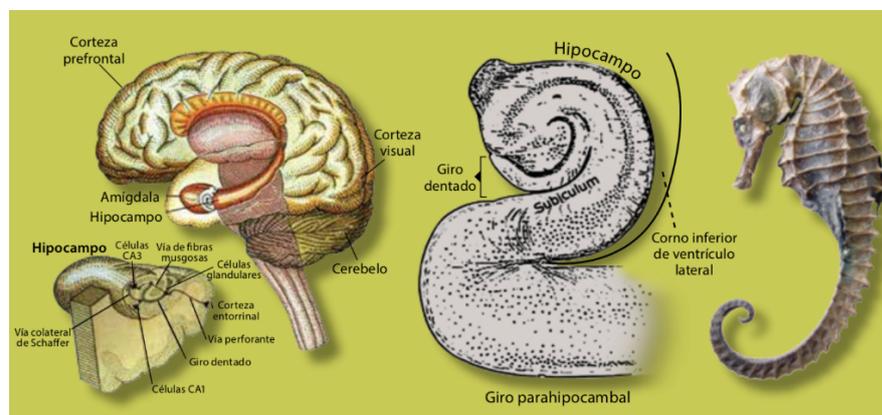


Figura 1. De izquierda a derecha: localización del hipocampo y la corteza entorrinal en el cerebro, corte coronal del lóbulo temporal medial donde se observa la forma del hipocampo y figura de un caballito de mar, de donde obtiene su nombre debido a su similitud (Aguilar Roblero, 2015).

En cuanto al tamaño del hipocampo, se conoce que oscila aproximadamente entre 3,5 y 4 cm de longitud anteroposterior, aunque este puede variar en función de las capacidades que desarrolle el sujeto, por ejemplo, se conoce que el tamaño del hipocampo aumenta en personas que desarrollan mecanismos de orientación espacial complejos, como prueba de ello, en el año 2000 la doctora E. Maguire realizó un estudio en taxistas londinenses, donde se confirmó que estos contaban con un hipocampo que crecía a medida que estos memorizaban nuevas calles. A pesar de su pequeño tamaño, desempeña un papel fundamental en la función cerebral, pues es responsable de la memoria reciente, la memoria remota y de la interpretación del desplazamiento espacial a la vez que se encuentra estrechamente comunicado con otras zonas del cerebro.

El hipocampo tiene un papel importante en la memoria episódica y espacial, de modo que involucra la capacidad de almacenar, aprender y recuperar información sobre experiencias de la vida (Squire, 1992). Conocer cómo funcionan los recuerdos del desarrollo de la vida de un sujeto, puede dar lugar a entender la naturaleza de los cambios que se producen en pacientes con enfermedades neurodegenerativas, como es el caso de la enfermedad de Alzheimer, donde tanto el hipocampo como la corteza entorrinal se encuentran afectados, provocando pérdida de la memoria y desorientación espacial.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Concretamente en pacientes con enfermedad de Alzheimer el hipocampo se observa más atrofiado. Esta pérdida de volumen es consecuencia de la muerte neuronal que desencadena este tipo de demencia. En la Figura 2 se observa la imagen en resonancia magnética del hipocampo de un paciente sano, un paciente con deterioro cognitivo leve temprano y un paciente con Alzheimer.

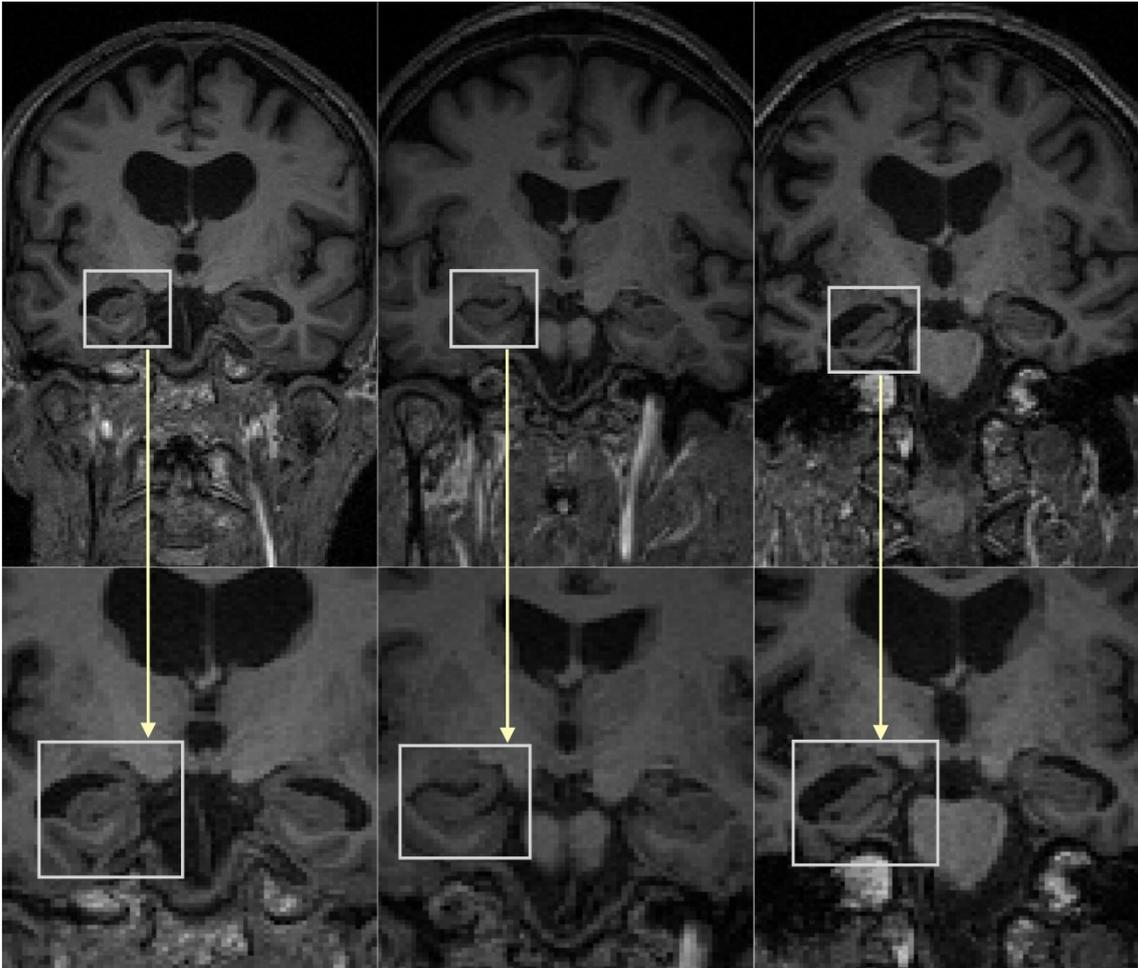


Figura 2. De izquierda a derecha: diferencia hipocampo entre paciente control, deterioro cognitivo leve y paciente con enfermedad de Alzheimer. Fuente: elaboración propia.

2.3 Neuroimagen

Las técnicas de neuroimagen son una modalidad de imagen médica que permite la visualización no invasiva del sistema nervioso central (SNC), en concreto del cerebro. Existen numerosas pruebas de neuroimagen tanto funcional como estructural, que aportan información esencial para el diagnóstico y el tratamiento de pacientes con trastornos que afecten al SNC, como es el caso de la enfermedad de Alzheimer. Las principales técnicas de neuroimagen empleadas son la tomografía axial computerizada (TAC o CT por sus siglas en inglés) y la resonancia magnética nuclear (RMN o RM), aunque existen otras técnicas de neuroimagen que destacan debido a que permiten estudiar la funcionalidad cerebral, como son la tomografía por emisión de positrones (PET) o la tomografía por emisión de fotón único (SPECT).

2.3.1 Imagen por Resonancia Magnética

La resonancia magnética o resonancia magnética nuclear (RMN) es una técnica de imagen médica no invasiva y no ionizante que permite la observación de tejidos. Desde que esta técnica se introdujo en la medicina clínica, ha demostrado ser una herramienta efectiva en el estudio del cerebro humano, pues proporciona información completa y de buena resolución sobre los tejidos blandos, lo que permitió la mejora de la calidad del diagnóstico y el tratamiento de patologías cerebrales (Chaplot, Patnaik y Jagannathan, 2006).

Las imágenes de resonancia magnética (IRM o MRI en inglés) proporcionan información bidimensional y tridimensional de órganos y estructuras dentro del cuerpo. La ventaja más importante de la RM es que es una técnica no invasiva. El nivel de resolución de las imágenes obtenidas mediante resonancia magnética destaca notablemente en comparación con otras modalidades de imágenes como las imágenes obtenidas por TAC.

El principio físico que emplea la resonancia magnética se basa en la presencia de átomos en los tejidos con un número impar de protones y/o neutrones en su núcleo (Martí-Climent *et al.*, 2010). Estas, al encontrarse bajo un campo magnético, absorben energía, excitándose y orientándose con una determinada dirección. Posteriormente, tras cesar la excitación, tardarán un tiempo determinado en relajarse y volver a su estado de equilibrio durante el cual emitirán energía, este tiempo dependerá de las propiedades del tejido en el que se encuentren. La energía liberada por los protones es recogida por un receptor para posteriormente ser analizada por un ordenador encargado de transformar dicha información en imágenes.

Desde que el físico Isidor Isaac Rabi describiera el fenómeno de la resonancia magnética en 1928 los equipos de resonancia magnética han evolucionado notablemente. En la actualidad existe una amplia variedad de equipos de resonancia magnética, sin embargo, su funcionamiento a la hora de adquirir y almacenar las imágenes es genérico, por lo que todos coinciden en que requieren de varios elementos esenciales para la obtención de imágenes:

En primer lugar, encontramos el imán, cuya principal función es emitir un campo electromagnético uniforme de elevada y constante intensidad, bajo el cual estará sometido el paciente. La morfología de los imanes de uso clínico puede variar en función de la estructura que deseamos observar, al igual que su potencia, la cual suele oscilar entre 1,5 o 3 Teslas (T). A su vez, encontramos un total de tres bobinas generadoras de gradientes de campo magnético, cada una

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

de ellas será responsable de hacer variar el campo magnético en cada una de las tres direcciones espaciales (transversal, frontal y sagital). Por último, contaremos con una bobina de radiofrecuencia que actuará como antena emisora de pulsos, estos excitarán las moléculas que se encuentren bajo la acción del campo magnético uniforme, es decir, las moléculas de los tejidos del paciente. A la combinación de pulsos de radiofrecuencia que se aplica se le denomina secuencia. Para acentuar la diferenciación entre tejidos, los parámetros físicos y los tiempos de cada secuencia de pulsos deben ser ajustados, de modo que se potencia la imagen en tres modalidades básicas: densidad protónica, T1 o T2. Una vez cese la excitación, las moléculas se relajarán, emitiendo un eco que será recogido por las antenas receptoras, tal y como se observa en la Figura 3, para posteriormente ser almacenado como datos numéricos y a continuación ser decodificado para traducir dichos datos en la imagen médica final.

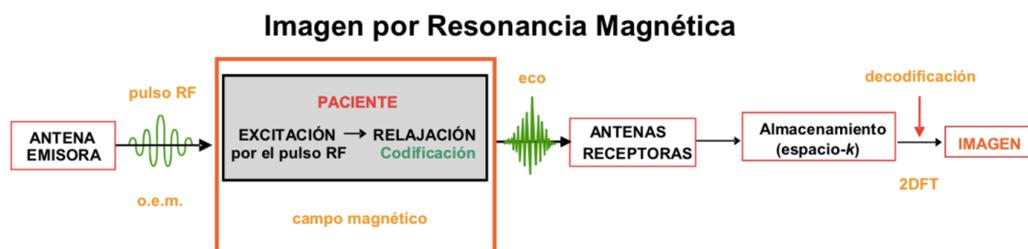


Figura 3. Esquema del procedimiento de obtención de una imagen por resonancia magnética (Moratal, 2017)

Las imágenes de resonancia magnética constituyen una herramienta de especial interés en el diagnóstico de patologías que afecten a las estructuras cerebrales, no sólo porque permiten estudiar el cerebro a partir de su visualización, sino que a su vez permiten extraer características para realizar un análisis cuantitativo de la imagen. La extracción de características de las IRM cerebrales es imprescindible para el análisis adecuado de estas imágenes. Estas características pueden ser estudiadas empleando métodos estadísticos o matemáticos, o incluso métodos de aprendizaje máquina.

2.3.2 Biomarcadores

Un biomarcador es una característica medible y evaluable capaz de indicar de forma objetiva si un proceso biológico es normal o patológico.

A pesar de los numerosos estudios que tratan de mejorar las técnicas de detección precoz de Alzheimer, como ya se ha comentado, los criterios actuales existentes no permiten un diagnóstico temprano ni etiológico, es por ello que, en los últimos años han surgido numerosos estudios que tratan de encontrar biomarcadores para la detección y el análisis de esta enfermedad (Valls-Pedret, Molinuevo y Rami, 2010). Contando con criterios de diagnóstico que se sustenten sobre biomarcadores objetivos se podría evaluar el transcurso de la enfermedad de manera más detallada y lograr predecir así la respuesta clínica a un posible tratamiento.

Se ha demostrado como el uso de biomarcadores como la determinación de proteínas implicadas en la LCR, las mediciones del hipocampo y el córtex entorrinal por resonancia magnética y la tomografía por emisión de positrones (tanto de glucosa como de amiloide), solos o en

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

combinación, pueden hacer posible efectuar diagnósticos más tempranos y etiológicos de la enfermedad de Alzheimer (Molinuevo Guix, 2018).

2.3.3 Análisis de texturas

La textura de una imagen se refiere a la apariencia, estructura y disposición de las partes de un objeto que pertenece a una imagen (Castellano *et al.*, 2004). Concretamente en imágenes digitales, este concepto se refiere a la distribución de valores de nivel de gris entre los píxeles de una región de interés (ROI, *Region Of Interest*) dada en la imagen 2D o en los vóxeles de un volumen de interés (VOI, *volume of interest*) de la imagen 3D. Con el fin de facilitar el análisis de los datos, todos los resultados obtenidos a lo largo de este trabajo tendrán la etiqueta de ROI a pesar de tratarse de un volumen, y se especificará en todo momento si se trata de una región circular (2D) o esférica (3D).

El concepto de ROI se puede extrapolar al estudio de imágenes diagnósticas. Por ejemplo, en el caso de la enfermedad de Alzheimer, como ya se ha comentado, el hipocampo es una de las principales regiones cerebrales que se ven afectadas durante el transcurso de la enfermedad, si seleccionamos esta zona como región de interés, sería posible realizar un análisis de texturas con el fin de obtener información de las imágenes médicas no apreciable a simple vista.

Se trata de un campo de investigación en curso, con aplicaciones que van desde la segmentación de estructuras anatómicas específicas y la detección de lesiones, hasta la diferenciación entre tejido patológico y sano en diferentes órganos.

El análisis de texturas implica un conjunto de cálculos matemáticos realizados con los datos contenidos en las imágenes, concretamente evalúa la posición e intensidad de los niveles de grises de los píxeles dentro de la ROI. De hecho, las características de textura (*texture features*) son los parámetros matemáticos que se calculan a partir de la distribución de píxeles, que caracterizan el tipo de textura y, por lo tanto, la estructura subyacente de los objetos a los que pertenecen los píxeles en la imagen.

Existen múltiples métodos para evaluar las características de textura y las interrelaciones de los píxeles, estos se pueden clasificar en cuatro grupos (Castellano *et al.*, 2004):

- Métodos estructurales
- Métodos basados en modelos
- Métodos estadísticos
- Métodos de transformación

Cada uno de estos métodos evalúa las interrelaciones entre píxeles teniendo en cuenta distintos parámetros. Los métodos estructurales estudian una serie de estructuras consideradas como básicas o primitivas que definen las texturas. Por otra parte, los métodos basados en modelos tratan de representar las texturas utilizando modelos matemáticos, mientras que los métodos de transformación aplican transformadas como Fourier o Wavelet con el fin de representar la imagen en un nuevo espacio donde ser estudiada. Por último, en cuanto a los métodos estadísticos, estos se centran en estudiar las distribuciones espaciales de los niveles de grises.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Para llevar a cabo el análisis de texturas en este proyecto se utilizarán tres parámetros que destacan dentro de este último grupo: el histograma de primer orden, la matriz de largo recorrido (GLRLM, *Gray-Level Run Length Matrix*) y la matriz de co-ocurrencia (GLCM, *Gray-Level Co-occurrence Matrix*) (Castellano *et al.*, 2004). El histograma de primer orden es un parámetro estadístico que no tiene en cuenta las interacciones espaciales entre píxeles pertenecientes a la ROI, únicamente describe características relacionadas con la distribución de los niveles de grises, de modo que su principal finalidad es plasmar de forma gráfica la suma del número de píxeles que representan un determinado nivel de gris. Por otro lado, tanto la GLRLM como la GLCM sí que consideran las interacciones entre píxeles de la ROI y, como consecuencia, para poder elaborar dichas matrices será necesario definir previamente una dirección en la que se evalúen los niveles de grises de la imagen. La matriz de largo recorrido representa el número de píxeles consecutivos que presentan el mismo nivel de gris en la dirección seleccionada. Por otro lado, para calcular la matriz de co-ocurrencia será necesario definir también un espacio de separación entre píxeles, de modo que esta matriz represente la frecuencia con la que se relacionan determinados niveles de grises, por ejemplo, definiendo cuántos píxeles con un valor de gris determinado se encuentran separados por una distancia concreta en la dirección seleccionada.

El potencial del análisis de texturas reside en que, gracias a que permite la extracción de una gran cantidad de información a partir de una imagen de resonancia magnética, combinado con técnicas de inteligencia artificial como el aprendizaje automático, las texturas pueden llegar a conformar un biomarcador fiable que facilite el diagnóstico precoz de la enfermedad del Alzheimer y el seguimiento de su progresión (Zhang *et al.*, 2012).

2.4 Métodos de aprendizaje máquina.

El aprendizaje máquina, aprendizaje automático o *machine learning* (ML) es una rama de la inteligencia artificial que trata de desarrollar sistemas capaces de optimizar un criterio de desempeño usando datos de experiencias pasadas.

La cantidad de datos de las que disponemos en la actualidad es demasiado extensa para su interpretación manual y, por lo tanto, existe una amplia necesidad de desarrollar herramientas de análisis de imágenes automatizadas. El aprendizaje automático tiene un elevado potencial en este ámbito, pues combina áreas como la estadística y otras ciencias matemáticas con el fin de cuantificar e interpretar dichos datos de forma objetiva.

El aprendizaje máquina destaca por ser una herramienta aplicable a múltiples campos. En función del enfoque, se puede dividir en:

- Aprendizaje memorístico.
- Aprendizaje basado en analogías.
- Aprendizaje inductivo.
- Aprendizaje por refuerzo.

Entre todos estos enfoques, se destacará el inductivo, pues este se basa en ejemplos específicos para extrapolar conclusiones y obtener un modelo que represente dicho comportamiento. Dentro de este enfoque, encontramos modelos de aprendizaje supervisado o discriminativo y no supervisado o generativo, en función de si existe una serie de datos previos que puedan ser usados para el entrenamiento del sistema o, por el contrario, se deben emplear funciones de probabilidad para llevar a cabo la clasificación. A su vez, el enfoque inductivo es utilizado, entre otras cosas, para la selección de características y la reducción de la dimensionalidad, aunque su interés reside especialmente en la elaboración de modelos predictivos que permitan la clasificación de distintas clases.

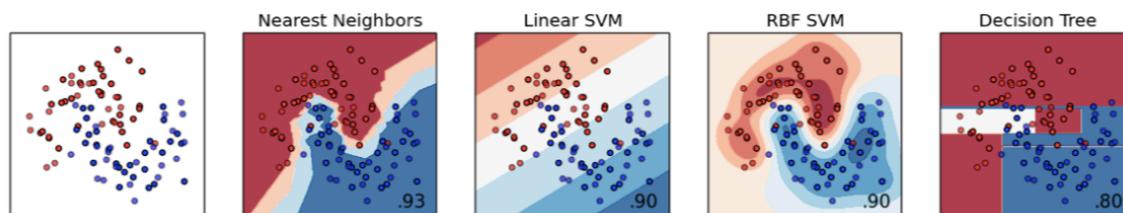


Figura 4. Tipos de función discriminante. Se observa que la clasificación de un conjunto de datos varía en función del método de clasificación de carácter discriminativo empleado. De izquierda a derecha se observa en primer lugar el conjunto de datos a clasificar, a continuación, las clasificaciones de dichos datos empleando los métodos de: K-vecinos más próximos, *Support Vector Machine* lineal, *Support Vector Machine* no lineal y árboles de decisión (*Classifier comparison*, fecha último acceso: 22-06-2018).

2.4.1 Métodos de clasificación

Los sistemas de aprendizaje máquina pueden realizar diversas tareas, entre las cuales destaca la predicción, que puede utilizarse con el fin de la estimación de resultados o la clasificación de datos. Concretamente, un clasificador o modelo de clasificación es una función que trata de evaluar el grado de relación entre un conjunto de instancias que pertenecen al mismo tipo o clase.

Previamente a aplicar cualquier método de clasificación es necesario realizar una partición del conjunto total de datos en dos subgrupos que serán utilizados con el fin de entrenamiento (*training*) y prueba (*test*), respectivamente. Dicha partición puede realizarse de forma aleatoria o controlada. El subconjunto de datos de entrenamiento se utiliza principalmente con el fin de estimar los parámetros del modelo, mientras que los datos de test son usados para comprobar el comportamiento del modelo estimado.

En la actualidad existe un amplio estado del arte que trata de englobar los métodos de aprendizaje automático que se utilizan para evaluar la influencia de las diferentes herramientas de procesamiento de imágenes en la precisión predictiva. Los detalles de implementación de algunos de los métodos clasificatorios más comúnmente empleados serán discutidos en la siguiente sección.

2.4.1.1 *K-Nearest Neighbors*

El algoritmo de *k*-vecinos más próximos o *K-Nearest Neighbors* (KNN) se basa en un sistema de clasificación supervisado, donde a partir de un conjunto de datos que se encuentran en un espacio dimensional, realiza una clasificación basada en criterios de vecindad. Se trata de predecir una nueva muestra utilizando las *K* muestras más cercanas del conjunto de entrenamiento, de modo que, para clasificar una un nuevo dato únicamente se tiene en cuenta la distancia a sus *K* vecinos más próximos y a qué grupo pertenecen, la clase pronosticada para la nueva muestra será entonces la respuesta con la estimación de probabilidad más alta calculada (Kuhn y Johnson, 2013).

La metodología de este modelo se basa en que, en primer lugar, durante la fase de entrenamiento, se almacenan los vectores característicos y las etiquetas de las clases a partir de un conjunto determinado de datos (*training*). A continuación, se llevará a cabo la fase de prueba, donde una nueva muestra de la que desconocemos su clase (*test*), será representada por un vector en el espacio característico. Se calculará la distancia entre los *K* vectores previamente almacenados durante el entrenamiento y el vector de la nueva muestra y se seleccionarán los ejemplos más cercanos, de modo que la muestra será clasificada en la clase donde haya sido más veces clasificada.

Fundamentalmente este método depende de la distancia entre muestras. Para calcularla, existen múltiples fórmulas, entre las cuales destacan la distancia euclídea, Minkowski, Manhattan, Chebyshev y Mahalanobis, cuyas ecuaciones se desarrollan a continuación.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Teniendo en cuenta que la distancia entre dos puntos x e y pertenecientes al espacio dimensional viene dada como $d(x, y)$, se define:

- Distancia de Minkowski

$$d(x, y) = \left(\sum_{j=1}^J |x_j - y_j|^r \right)^{\frac{1}{r}} \quad r \geq 1 \quad (1)$$

Donde r es el orden.

- Distancia Euclídea

$$d(x, y) = \sqrt{\sum_{j=1}^J (x_j - y_j)^2} \quad (2)$$

- Distancia de Manhattan o *Cityblock*

$$d(x, y) = \sum_{j=1}^J |x_j - y_j| \quad (3)$$

- Distancia de Chebyshev

$$d(x, y) = \max_{j=1..J} |x_j - y_j| \quad (4)$$

La elección del número de vecinos depende fundamentalmente de las características de los datos, pues se conoce que valores pequeños de K crean muchas pequeñas regiones de cada clase, mientras que altos valores de K conducen a la formación de menos regiones de mayor tamaño, reduciendo el efecto de ruido en la clasificación (ver Figura 5). Además, las funciones son discriminantes a trozos, pues las fronteras de decisión se encuentran únicamente determinadas por conjunto de puntos (muestras).

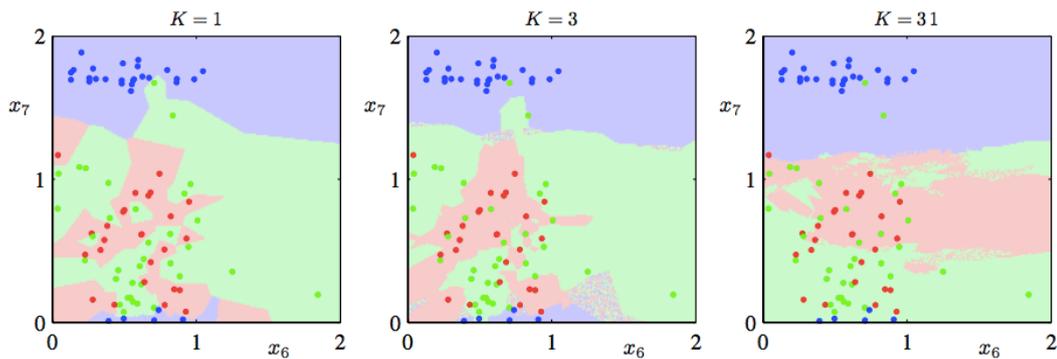


Figura 5. Ilustración de las fronteras y regiones definidas por modelos K-vecinos con distintos valores de K . Se observa que, a medida que K crece, las fronteras de decisión son menos abruptas (Bishop, 2013).

2.4.1.2 Support Vector Machine

El método de máquinas de vectores de soporte o *Support Vector Machine (SVM)* es un método de modelado de carácter discriminativo empleado principalmente para llevar a cabo clasificaciones lineales y binarias, aunque puede extenderse a clasificaciones no lineales.

La formulación de este método se basa en generar hiperplanos definiendo fronteras lineales de separación que particionen el espacio de características según el problema de dos clases. El hiperplano que esté más separado de ambos conjuntos de muestras, es decir, que tenga un mayor margen, será la solución. Los vectores de soporte (*support vectors*) serán las muestras más cercanas al hiperplano de cada clase. Para encontrar la frontera y calcular el hiperplano, el sistema se entrena con un número limitado de muestras del conjunto de entrenamiento que cumplen con una serie de características denominadas vectores de soporte. Si imaginamos un plano definido por dos clases separadas en función de sus características (Figura 6), cada una de las clases tendrá un vector de soporte.

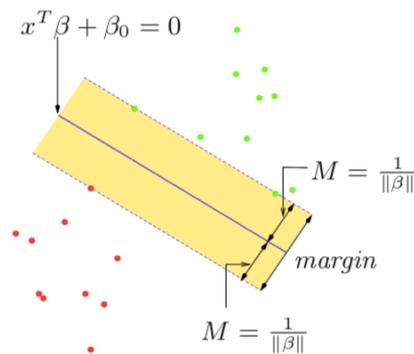


Figura 6. Clasificación con SVM. En rojo y verde se observan las dos clases. Los puntos que se encuentran sobre las líneas discontinuas se corresponden con los vectores de soporte. El plano de esta última región intermedia a estas dos líneas es la solución del problema (Hastie, Tibsharani y Friedman, 2009).

En ocasiones será necesario añadir determinadas restricciones al modelo con el fin de evitar sobreajustes en la clasificación. Por ejemplo, en el caso de que la muestra de datos no sea perfectamente separable, el algoritmo no será tan restrictivo (tendrá un coste bajo) y permitirá algunas clasificaciones erróneas con la finalidad de poder definir el hiperplano, por el contrario, un mayor coste, significará que los puntos mal clasificados tendrán más peso y más estricta será la separación de los datos.

Por otra parte, también cabe mencionar que la ecuación original de decisión para llevar a cabo la clasificación de una nueva muestra u desconocida establece fronteras de separación lineales:

$$D(u) = \beta_0 + \sum_{i=1}^n y_i \alpha_i x_i' u \quad (5)$$

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Sin embargo, existirán casos en los que las clases no serán separables linealmente, sino que será necesario establecer otro tipo de límites de clasificación. Para ello, será necesario realizar una transformación del espacio original a un espacio de mayor dimensión, esta tarea se lleva a cabo introduciendo en la ecuación original una función denominada *kernel*, $K(x_i, u)$:

$$D(u) = \beta_0 + \sum_{i=1}^n y_i \alpha_i K(x_i, u) \quad (6)$$

En casos lineales, $K(x_i, u)$ será equivalente a $x_i' u$. Por el contrario, en casos no lineales, esta función puede ser sustituida por transformaciones de tipo polinómico, radial o hiperbólico (Kuhn y Johnson, 2013). De este modo se consigue que el modelo SVM produzca límites de decisión extremadamente flexibles que se adapten de la mejor forma posible a la clasificación del conjunto de datos.

Cabe destacar que la elección de los parámetros de la función kernel y el valor de costo deben ser ajustados de forma correcta para que el modelo no sobrepase los datos de entrenamiento. Si el valor de costo es bajo, el modelo puede subestimar los datos, por el contrario, cuando este es alto, los datos pueden ser sobreajustados, especialmente si el valor del kernel es también elevado.

El concepto de la separación de hiperplanos sobre el cual trabaja el método de SVM no es aplicable en el caso de que se requiera la clasificación de más de dos clases. En este caso, para realizar clasificaciones multiclase será necesario introducir los conceptos de *one-vs.-one* o uno contra uno y *one-vs.-all* o uno contra todos (Casella, Fienberg y Olkin, 2006):

- One vs. one. Para esta clasificación se tendrá un total de $\frac{K*(K-1)}{2}$ clasificadores binarios, donde K es el número de clases (por ejemplo, para tres clases tenemos tres clasificadores binarios). Cada uno de ellos, recibe muestras de un par de clases distintas del conjunto de datos de entrenamiento y debe aprender a distinguir entre esas dos clases. Cuando el modelo debe clasificar una nueva muestra de test, todos los clasificadores tratan de clasificarla. Posteriormente se cuenta el número de veces que dicha muestra ha sido asignada a cada una de las K clases y la muestra pasa a pertenecer a la clase en la que fue más frecuentemente clasificada.
- One vs. all. En este caso se entrena un único clasificador por clase, de modo que se compara cada una de las K clases con las clases K-1 restantes, es decir, el modelo aprende a distinguir una clase de todas las demás. Cuando llega una nueva muestra para ser clasificada, el modelo calcula determinados parámetros con el fin de comprobar si la nueva muestra pertenece a la clase que este conoce.

2.4.1.3 Decision Trees

Los árboles de decisión o *Decision Trees* (DT) representan un método que clasifica los datos según un conjunto de reglas o condiciones. Estas estructuras están formadas por nodos, ramas y hojas, donde cada nodo representa una característica o variable, cada rama un posible valor para dicha característica y cada hoja representa los posibles valores finales.

Se trata de una técnica muy intuitiva y fácil de comprender gracias al carácter descriptivo de la segmentación. Para construir dichas estructuras jerárquicas, se parte de una muestra de datos de entrenamiento que servirá para determinar el criterio de clasificación, de este modo, se generará el primer nodo del árbol en base a la variable escogida para separar los datos y su umbral óptimo, si estos cumplen la condición, serán clasificados en un grupo, si no, en otro. Estos pasos se repiten hasta que se genera el árbol final y los datos quedan correctamente clasificados en distintas clases.

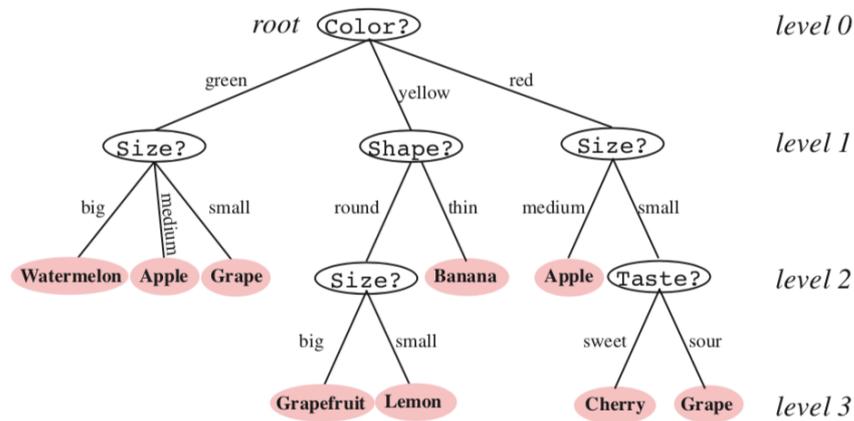


Figura 7. Ejemplo de aplicación de árbol de decisión top-down para clasificar frutas (hojas) en función de características como el tamaño o el color (nodos). El procedimiento se basa en ir visitando nodos sucesivamente hasta llegar al nodo hoja, donde se lee la etiqueta de la categoría (Duda, Hart y Stork, 2001).

Los algoritmos para la construcción de árboles de decisión suelen trabajar de manera *top-down* (Figura 7), es decir, partiendo de una generalización que se refine a formas más específicas con cada separación, escogiendo en cada paso la variable que mejor divide el conjunto de elementos.

Gran parte del trabajo en el diseño de árboles de decisión se centra en decidir qué consulta se debe realizar en cada nodo para obtener la mejor clasificación. El principio fundamental en el que se resume este diseño es la simplicidad: son preferibles las decisiones que conducen a un árbol simple y compacto con pocos nodos, de modo que se garantice que los datos lleguen a los últimos nodos o nodos hoja tan “puros” como sea posible (Duda, Hart y Stork, 2001).

Para evaluar la pureza de la clasificación, se usan diferentes criterios. Entre ellos, destaca el índice Gini, el cual se calcula para cada nodo t :

$$GINI(t) = 1 - \sum_j [p(j|t)]^2 \quad (7)$$

donde $p(j|t)$ es la frecuencia relativa con la que una muestra j es clasificada según el nodo t .

Los índices más cercanos a cero indicarán una mayor pureza en la separación, por el contrario, los más cercanos a 1 indicarán mayor impureza. Por ejemplo, si tenemos un total de seis datos que, tras el primer nodo son clasificados en dos clases: dos datos en la primera clase y cuatro en la segunda, se calcularía una frecuencia relativa de $(2/6)^2$ y $(4/6)^2$. El índice de GINI para dicho nodo sería igual a 0,444.

Los árboles de decisión han emergido como método de aprendizaje gracias a que poseen propiedades que los hacen destacar frente a otros métodos, por ejemplo, aportan ventajas como la facilidad para su construcción e interpretación. Sin embargo, cabe destacar que esta herramienta no garantiza siempre resultados exactos, pues estos dependen fundamentalmente de la complejidad del diseño del árbol y del tamaño muestral (Hastie, Tibsharani y Friedman, 2009).

2.4.2 Métodos de selección de características

Uno de los principales problemas a los que se enfrenta el aprendizaje máquina es la estimación de la calidad de las características para realizar la clasificación ya que, en la mayoría de los problemas de aprendizaje automático existen cientos o miles de posibles características que describen cada objeto de entrada. Estas circunstancias provocan que el comportamiento de los métodos de clasificación no sea el más idóneo, debido a que, aumentan significativamente su complejidad y el tiempo y espacio necesario para el procesamiento de los datos. Además, desde un punto de vista estadístico, un exceso de características irrelevantes puede provocar ruido en el modelo y, como consecuencia, la pérdida de información.

Una forma común de resolver esta situación es la selección de características o *feature selection*. Los métodos de *feature selection* son técnicas que se emplean para determinar cuáles son las características más relevantes a la hora de realizar una clasificación. Estos parten de un conjunto de características de entrada, de modo que estas son ordenadas en función de su relevancia elaborando un ranking. A partir de dicho ranking, se selecciona un subconjunto de las n primeras características consideradas suficientemente representativas, de forma que se garantiza que el modelo clasificatorio funcione de forma eficiente.

Existen numerosos algoritmos que tratan de evaluar la calidad de las características que forman parte de un modelo clasificador. En términos generales, estos pueden clasificarse en tres grandes familias: métodos de filtro (*filter*), métodos *wrapper* y métodos embebidos (*embedded*). Los métodos de filtro clasifican las características antes de aplicar el algoritmo de aprendizaje seleccionando aquellas características con puntuaciones o *scores* de alto rango. Los métodos *wrapper* califican las características usando las puntuaciones obtenidas al aplicar un algoritmo de aprendizaje. Por último, los métodos *wrapper* combinan la selección de características con el algoritmo de aprendizaje y su diseño está estrechamente vinculado con un algoritmo de aprendizaje específico, que a su vez limita su aplicación a otros algoritmos de aprendizaje.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

En este trabajo, se hará especial hincapié en los métodos de filtro para la selección supervisada de características. Estos cuentan con la ventaja de que tienden a ser simples y rápidos, además de robustos frente al sobreajuste (*overfitting*), aunque su procedimiento, por lo general, tiene una naturaleza subjetiva debido a que la mayoría de los métodos de puntuación de características no tienen un punto de corte obvio para declarar qué predictores son lo suficientemente importantes como para formar parte del modelo. Además, también cabe mencionar que las características que estos seleccionan no siempre son óptimas. Por un lado, dado que la puntuación de cada característica es calculada independientemente, se descuida la combinación de características, lo que significa que no evalúa dos o más características juntas. Por ejemplo, podría ser el caso de que los *scores* o puntuaciones de dos características por separado fueran bajos, pero el *score* de la combinación de ambas sea alto, en este caso, el método *filter* descartará ambas características. Por otro lado, esta técnica tampoco considera características redundantes, en el caso de que los *scores* de dos características sean altos y estén altamente correlacionados, este método selecciona ambas características, cuando una de ellas podría ser eliminada sin pérdida en el rendimiento, de hecho, se ha demostrado que eliminar características redundantes puede mejorar el rendimiento del método de clasificación (Gu, Li y Han, 2012).

A pesar de dichas desventajas, los métodos de *filter* son comúnmente usados, pues logran eliminar los atributos menos interesantes para el modelo clasificatorio, a la vez que otorgan buenas características computacionales y de robustez frente al *overfitting*. Es por ello que, los métodos de *feature selection* empleados para la elaboración de este trabajo serán exclusivamente tipo *filter*, concretamente nos referiremos a métodos basados en el P-valor, Fisher, Relief y Relief-F.

2.4.2.1 P-valor

Dentro de los métodos *filter* para la selección de características es común emplear métodos estadísticos como pruebas de hipótesis (por ejemplo, *t*-test). Estos métodos generan parámetros como el *p*-valor con el fin de determinar qué predictores tienen relaciones estadísticamente significativas con el resultado. Para ello, se calculan niveles de confianza (α) que se usan como límite del umbral de confianza y permiten el cálculo del *p*-valor, de este modo se determina individualmente cuál es la tasa teórica de falso positivo para cada prueba.

El principal inconveniente reside en que, al calcular la tasa de falso positivo de forma individual, esta puede resultar pequeña, sin embargo, cuando se lleva a cabo este estudio con una gran cantidad de pruebas estadísticas simultáneamente, la probabilidad de falso positivo aumenta exponencialmente (problema de multiplicidad). Gracias a la modificación del *p*-valor de cada predictor, se pueden aplicar correcciones que resuelvan este problema, por ejemplo, el método de corrección de Bonferroni (Bland and Altman 1995) trata de garantizar que el *p*-valor del predictor sea menor a α/M , siendo M el número de pruebas realizadas, de modo que se controla la probabilidad de resultados falsos positivos (Kuhn y Johnson, 2013).

De este modo, el método de tipo filtro basado en *p*-valor se centra en restringir los predictores con *p*-valor inferior a un determinado número, es decir, acepta una determinada tasa de falso positivo para cada comparación individual y, posteriormente, aplica algún tipo de corrección para evitar el problema de la multiplicidad, de modo que los predictores se detengan cuando sus *p*-valores sean menores que $\frac{\alpha}{\text{número de pruebas}}$.

2.4.2.2 Fisher

El análisis discriminante de Fisher es uno de los métodos de selección de características más comúnmente utilizado, y fue desarrollado por R.A Fisher en 1936 (Fisher, 1936). Este se basa en la idea de que las características representativas para la clasificación deben asignar valores similares a las instancias de la misma clase y diferentes a las instancias de distintas clases (Tang, Alelyani y Liu, 2014). Siguiendo este principio, se calcula un *score* denominado *Fisher Score*, S_i , para evaluar cada característica:

$$S_i = \frac{\sum_{k=1}^K n_j (\mu_{ij} - \mu_i)^2}{\sum_{k=1}^K n_j \rho_{ij}^2} \quad (8)$$

donde μ_{ij} y ρ_{ij} son la media y la varianza de la característica i en la clase j respectivamente, n_j es el número de instancias en la clase j y μ_i es la media de la característica i .

Para encontrar la solución óptima a esta ecuación se aborda un enfoque heurístico que calcula una puntuación de forma individual para cada característica i , posteriormente, se elabora un ranking y seleccionan las características clasificadas con las puntuaciones más altas. Al tratarse de un método de tipo filtro, esta técnica cuenta con el inconveniente de que se evalúa individualmente cada característica. Recientemente se ha propuesto generalizar esta técnica para trabajar con subconjuntos de características (Gu, Li y Han, 2012).

2.4.2.3 Relief y Relief-F

El algoritmo *Relief* fue creado por Kira y Rendell en el año 1992 con el fin de descartar características irrelevantes en una clasificación (Robnik-Sikonja y Kononenko, 2003). Este utiliza puntos seleccionados al azar y sus vecinos más cercanos para evaluar cada predictor de forma aislada. Concretamente, su técnica se basa en realizar iteraciones a través de un procedimiento de actualización de peso: en primer lugar, selecciona aleatoriamente una muestra x del conjunto de entrenamiento, a continuación, ubica las muestras más cercanas a x de ambas clases, considerando muestras como aciertos o *hits* y errores o *misses*. Para cada predictor se calcula la diferencia entre los valores del predictor en el punto aleatorio y los aciertos y errores. El peso de cada característica es una acumulación de estas diferencias, de tal manera que aumenta si el acierto está cerca del valor seleccionado al azar y disminuye si el error está cerca, siendo los pesos más altos indicadores de la relevancia del predictor en el método (Kuhn y Johnson, 2013).

Kononenko mejoró este procedimiento en 1994 creando un nuevo algoritmo modificado a partir de *Relief*, llamado *ReliefF* o *Relief-F*. El método de *ReliefF* es un algoritmo que, a diferencia del *Relief*, permite tratar con más de dos clases y es más robusto, de modo que es capaz de trabajar con datos ruidosos o incluso incompletos mediante probabilidades. Al igual que el *Relief*, usa una métrica de diferencias, aunque emplea más de un vecino cercano. En primer lugar, *ReliefF* selecciona una instancia al azar y busca K de sus vecinos más cercanos de la misma clase y los bautiza como “aciertos cercanos”; seguidamente repite el mismo procedimiento buscando K vecinos más cercanos de clases distintas, llamándolos “errores más cercanos”, y finalmente, a partir de dicha información, actualiza la estimación del peso de las características (Robnik-Sikonja y Kononenko, 2003).

2.4.3 Validación de sistemas

Partiendo de la necesidad de dividir el conjunto de datos en dos subgrupos, uno de entrenamiento y otro de prueba para llevar a cabo técnicas de aprendizaje máquina, existen diversas estrategias de validación de modelos en función de cómo se haga dicha partición.

A continuación, se expondrán las características de algunas de las técnicas de validación de sistemas de aprendizaje más conocidas y que resultan de interés de acuerdo a las características de este trabajo.

2.4.3.1 Validación *hold-out*

El método de retención, método *hold-out* o validación simple es un método de validación implementado en 1979 por Devroye y Wagner que realiza una única división de la muestra de datos en dos subconjuntos: datos de entrenamiento o training y datos de prueba o test, generalmente siendo mayor el tamaño del subgrupo de entrenamiento (Arlot y Celisse, 2009). Los datos de entrenamiento se utilizan para llevar a cabo un análisis que posteriormente será validado con el subconjunto de datos de *test*, de modo que inicialmente se entrena al sistema con los datos de training, este elabora una función que calcula los valores de salida para dichos datos y, posteriormente, durante la fase de validación, el sistema analiza un nuevo conjunto de datos (*test*) empleando la función que previamente había elaborado.

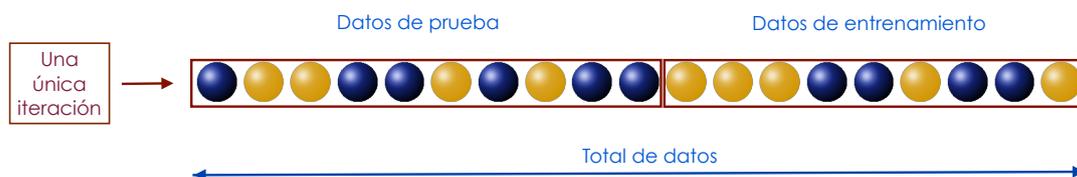


Figura 8. Método de validación *hold-out*. Se lleva a cabo una única partición del total de datos en dos subgrupos: datos de prueba o test y datos de entrenamiento o training, en una única iteración. Fuente: elaboración propia.

Se trata de un método de validación muy simple, cuya principal ventaja es la rapidez, sin embargo, no es un método preciso, pues los resultados varían enormemente en función de cómo sea la división entre datos de entrenamiento y prueba. Esto es debido al hecho de que sólo se utilizan una parte de las muestras disponibles para llevar a cabo el aprendizaje, lo que provoca la pérdida de información en el proceso de clasificación. Además, normalmente el modelo no se ajustará con igual exactitud a los datos de test que a los de entrenamiento, lo que significa que existe riesgo de *overfitting* o sobreajuste del modelo en caso de que el tamaño de los datos de entrenamiento sea pequeño o el número de parámetros demasiado grande.

El algoritmo debe ser capaz de predecir resultados a partir de lo aprendido a lo largo de su proceso de entrenamiento, de modo que la capacidad de resolver nuevas situaciones resulta un aspecto clave en cualquier método de aprendizaje. Es por ello que, el método *hold-out* fue mejorado, apareciendo así el concepto de validación cruzada.

2.4.3.2 Validación cruzada

La validación cruzada o *cross validation* (CV) es una técnica planteada por Geisser en 1975 que se emplea para evaluar análisis estadísticos, cuyos fundamentos se basan en el método de validación *hold-out*. Se plantea con el principal objetivo de evitar la pérdida de información que ocurría en el método de retención, la cual se debía a la pérdida de parte de los datos al llevar a cabo la partición en subgrupos de test y training.

En esta técnica, al igual que en el método de retención, los datos se dividen en dos subconjuntos (training y test), sin embargo, en este caso se realizarán diferentes divisiones de datos, tomando subgrupos de training y test distintos para cada iteración. Básicamente este método trata de calcular la media aritmética obtenida de las medidas de evaluación sobre cada partición en cada iteración (Arlot y Celisse, 2009). Además, la validación cruzada pretende optimizar los parámetros de ajuste del modelo de clasificación, de modo que este se adapte de la mejor forma posible a un nuevo conjunto de datos de test y se evite así el sobreajuste.

Existe una amplia variedad de métodos de validación cruzada, entre los cuales destacan (Arlot y Celisse, 2009): validación dejando uno fuera o *leave-one-out cross validation* (LOO-CV), validación dejando p datos fuera o *leave-p-out validation* (LPO), validación de K iteraciones o *K-Fold cross validation* y *bootstrap*.

De acuerdo con las características de este trabajo, a continuación, será explicada brevemente la metodología de actuación de la validación cruzada de K iteraciones o *K-fold cross validation*. Mediante esta técnica el conjunto de datos es dividido en K subconjuntos con aproximadamente el mismo número de muestras. En cada evaluación, se escogerán una serie de dichos subconjuntos ($K-1$) para realizar el entrenamiento, dejando siempre uno libre que será exclusivamente empleado para la fase de test. Los subconjuntos de prueba o test serán siempre independientes y distintos para cada iteración, mientras que los de entrenamiento no. El proceso se repite durante K iteraciones, con cada uno de los posibles datos de prueba. Finalmente, se obtiene un único resultado a partir del cálculo de la media aritmética de cada iteración.

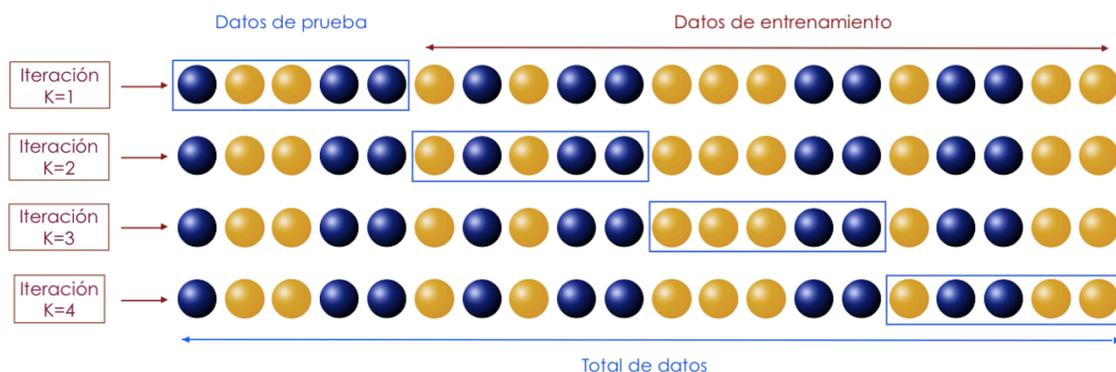


Figura 9. Método de validación cruzada con K iteraciones (en este caso $K=4$). Se observa en cada iteración existe una partición distinta de los datos, siendo siempre los datos de prueba nuevos para el sistema. Fuente: elaboración propia.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Este método destaca frente a la validación *hold-out* debido a su alta precisión, pues se evalúan K combinaciones distintas de datos y no una única, como ocurría en *hold-out*. Sin embargo, presenta el inconveniente de que es más lento, pues requiere de un coste computacional mayor, proporcional al número de iteraciones que requiera el proceso en cada caso.

Otros métodos alternativos a la validación cruzada de K iteraciones podrían ser el LOO-CV o el *bootstrap*. El LOO-CV, como su propio nombre indica, usa todos los datos de la muestra original para entrenar el modelo excepto uno, que se emplea posteriormente como muestra de prueba. Este procedimiento se repite hasta que todas las muestras hayan servido al menos una vez como conjunto de prueba. Se trata de un método muy costoso computacionalmente y destaca debido a que puede llevar a una estimación casi insesgada, sin embargo, esto ocurre a costa de una variabilidad que puede resultar muy significativa, es por ello que se determinó que para lograr un buen compromiso entre el sesgo y la varianza es preferible utilizar el método de K iteraciones con K=5 o K=10 (Lemm *et al.*, 2011).

Otra posible alternativa a implementar es el método de *bootstrap*, que es particularmente útil en caso de funciones discontinuas. Este procedimiento consiste en que se elabora una muestra o *bootstrap* a partir de un muestreo uniforme con reemplazo de la muestra original. Posteriormente, el *bootstrap* se utiliza para el entrenamiento, mientras que el conjunto omitido se utiliza para la validación. Sin embargo, este método tiene el inconveniente de que puede ser excesivamente optimista en modelos complejos debido a un sobreajuste de los datos.

Breiman y Spector llevaron a cabo en 1992 un estudio que comparó *leave-one-out cross validation*, *bootstrap* y *K-fold cross validation* en la selección de características, y este concluyó que, en lo que respecta al ámbito de selección de características, la validación *K-fold*, concretamente con K=10, superaba con creces a los otros métodos (Lemm *et al.*, 2011). Es por ello que, frente a las alternativas posibles, se decidió profundizar e implementar el método *K-fold* en el desarrollo de este trabajo.

2.4.4 Métodos de evaluación de resultados

Con el fin de evaluar la calidad de un modelo predictivo de clasificación y descartar modelos que no resulten lo suficientemente óptimos para llevar a cabo esta tarea, se emplean métodos como métricas o representaciones gráficas que facilitan la labor de interpretación de los resultados y la consiguiente tarea de identificación del mejor modelo clasificadorio.

Los métodos de evaluación de resultados principalmente se basan en el cálculo de ratios que posteriormente pueden ser interpretados empleando tablas de contingencia, más comúnmente conocidas como matrices de confusión, o, por el contrario, pueden ser interpretados a través de una representación gráfica como, por ejemplo, en forma de curvas ROC (*Receiver Operating Characteristic*).

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

2.4.4.1 Matriz de confusión

Dentro del campo del aprendizaje automático, concretamente en el contexto de un problema de predicción basado en dos únicas clases, donde una es positiva y la otra negativa, el concepto de verdaderos positivos (*true positives*, TP) hace referencia a las instancias clasificadas correctamente como procedentes de la clase positiva. Por otro lado, el concepto de falsos positivos (*false positives*, FP) se corresponde con las instancias clasificadas como clase positiva, pero que realmente pertenecen a la clase negativa. Los conceptos de verdaderos negativos (*true negatives*, TN) y falsos negativos (*false negatives*, FN) pueden ser explicados similarmente, pues se refieren a las muestras correcta e incorrectamente clasificadas como negativas, respectivamente.

La matriz de confusión es una herramienta que trata de plasmar estos conceptos de una manera visual, tal y como se muestra en la Tabla 2.

		<i>Predicción</i>	
		C_P	C_N
Clase real	C_P	Verdaderos Positivos (TP)	Falsos negativos (FN)
	C_N	Falsos positivos (FP)	Verdaderos negativos (TN)

Tabla 2. Estructura de matriz de confusión para un problema de clasificación binario (dos clases: C_P y C_N).
Fuente: elaboración propia.

Siguiendo este mismo razonamiento es posible emplear la matriz de confusión en clasificaciones de varias clases, tal y como se verá en el desarrollo de este proyecto, con el fin de analizar los resultados de las predicciones para pacientes con enfermedad de Alzheimer, pacientes sanos y pacientes con deterioro cognitivo leve.

A partir de dichos conceptos, y con el fin de evaluar la fiabilidad del modelo estimador, es posible calcular estadísticas que pueden ser relevantes cuando una clase es interpretada como evento de interés, en nuestro caso, la enfermedad de Alzheimer. Algunos de los conceptos estadísticos más empleados son la exactitud, la precisión, la sensibilidad o la especificidad.

La exactitud (*accuracy*) es posiblemente el parámetro más empleado en la evaluación de la precisión del modelo clasificador, este hace referencia a la tasa de elementos clasificados correctamente, y se calcula como:

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Sin embargo, cabe destacar que este parámetro no resulta del todo fiable en clasificaciones binarias donde las clases se encuentren desequilibradas, por ejemplo, donde tengamos 9990 ejemplos de la clase 1 pero únicamente 10 en la clase 2. En este caso el modelo de clasificación podría decir con una precisión del 9990/10000 (99,9%) que los ejemplos serían de la clase 1; esta afirmación sería engañosa pues nunca detectaríamos ejemplos de la clase 2.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Es por ello que, como alternativa se emplean métricas como la matriz de confusión u otras estadísticas que sean sensibles al coste y también puedan ser indicadoras de la fiabilidad del modelo clasificador. Una de ellas sería la precisión, la cual se define como el porcentaje de verdaderos positivos con respecto al total de positivos.

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (10)$$

Por otro lado, la sensibilidad, también conocida como *recall* o ratio de verdaderos positivos, muestra el porcentaje de muestras positivas correctamente identificadas:

$$\text{Sensibilidad (recall)} = \frac{TP}{P} = \frac{TP}{TP + FN} \quad (11)$$

mientras que la especificidad es un parámetro indicativo de cuantos casos negativos, se han dado como realmente negativos:

$$\text{Especificidad} = \frac{TN}{N} = \frac{TN}{TN + FP} \quad (12)$$

2.4.4.2 Curva ROC y AUC

En ocasiones será de especial interés disponer una única medida que refleje las tasas de falsos positivos y falsos negativos. El método más común para combinar la sensibilidad y la especificidad en un solo valor es la curva característica operativa del receptor (ROC) (Hastie, Tibsharani y Friedman, 2009).

Las curvas ROC se diseñaron como un método general que, dado un conjunto de puntos de datos continuos, determina un umbral efectivo tal que los valores por encima de dicho umbral sean indicativos de un evento específico. Estas curvas representan la sensibilidad o ratio de verdaderos positivos en su eje vertical frente al ratio de falsos positivos en el eje horizontal. Se trata de una herramienta útil a la hora de escoger un umbral que maximice de manera apropiada el equilibrio entre sensibilidad y especificidad. Inicialmente fueron definidas para problemas de dos clases, sin embargo, existen estudios que tratan de extender dicha herramienta para la comparación multiclases (Kuhn y Johnson, 2013)

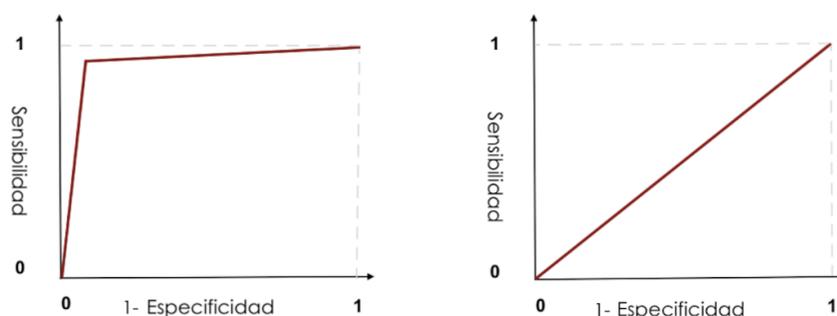


Figura 10. Ejemplo de curvas ROC. A la izquierda, una “buena clasificación”, con un AUC muy próximo a 1. A la derecha, una “mala clasificación”, con un AUC de 0.5. Fuente: elaboración propia

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Esta herramienta puede ser usada a su vez para evaluar cuantitativamente el modelo de clasificación binario, para ello se calcula el área debajo de dicho umbral (*Area Under the Curve*, AUC). Un modelo que separe completamente las dos clases, y que por lo tanto sea considerado como efectivo, tendría valores de especificidad y sensibilidad iguales a 1; gráficamente este hecho equivaldría a una curva con forma de un único escalón cuya altura crezca desde las coordenadas (0,0) hasta (0,1) para posteriormente mantenerse constante desde (0,1) a (1,1). De este modo, el área bajo la curva (AUC) para dicho modelo sería igual a 1. Por el contrario, un modelo ineficaz daría como resultado una curva ROC que siga de cerca la diagonal de 45° trazada como referencia, teniendo un valor de AUC que se aproxima a 0,5, tal y como se aprecia en la Figura 10.

Una ventaja de usar curvas ROC para caracterizar modelos es que, la curva es insensible a las disparidades en las proporciones de cada clase, debido a que únicamente es función de la sensibilidad y la especificidad (Duda, Hart y Stork, 2001). Sin embargo, usar exclusivamente el parámetro AUC para evaluar modelos puede conducir a errores ya que, por ejemplo, un modelo puede producir una pendiente pronunciada de curva ROC a la izquierda, pero tiene un AUC menor que otro modelo, en este caso debería primar la posición de la curva frente al valor del área, pues el valor de AUC no identificaría el mejor modelo (Kuhn y Johnson, 2013).

3. MATERIALES

A lo largo de esta sección se detallarán cuáles han sido los materiales empleados para el desarrollo de este estudio. Se describirá el conjunto de datos obtenidos a partir de las IRM que se utilizan como materia prima en este estudio. También se detallan las propiedades y los detalles de implementación de los métodos de aprendizaje automático que se utilizan para la clasificación binaria, todos implementados con el programa MATLAB.

3.1 Datos

El conjunto de datos o *data set* es la base de todo sistema de predicción, pues se corresponde con los elementos que hemos empleado para entrenar y testear nuestro modelo. Concretamente el conjunto de datos empleado para este estudio consiste en 150 imágenes cerebrales obtenidas mediante técnicas de resonancia magnética potenciadas en T1. Estas imágenes se encuentran en formato NIFTI y provienen de tres grupos distintos de pacientes: sujetos control o sanos, sujetos con deterioro cognitivo leve temprano y sujetos con enfermedad de Alzheimer. Dichas imágenes fueron adquiridas usando un equipo de 3T (Siemens Trio) con el siguiente protocolo: secuencia de adquisición 3D MP-RAGE, tiempo de repetición (TR) / tiempo de eco (TE) = 2300/2.98ms, ángulo de inclinación de 9 grados, resolución “in-plane” de 1x1 mm² siendo el plano de adquisición el sagital, matriz de adquisición de 256x256 y campo de visión de 256 mm.

En total, se han empleado datos de 50 sujetos control (*Control*, CN), 50 sujetos con deterioro cognitivo leve temprano (*Early Mild Cognitive Impairment*, EMCI) y 50 sujetos con Alzheimer (*Alzheimer Disease*, AD). El término de EMCI surge como consecuencia de la necesidad de clasificar los distintos estadios de progresión de la demencia, concretamente esta fase forma parte de las etapas más tempranas del deterioro cognitivo leve, de acuerdo a la escala GDS, este tipo de deterioro cognitivo estaría relacionado con la fase 2 (ver Tabla 1). Concretamente estos pacientes fueron seleccionados para este proyecto con el fin de evaluar las posibles fases tempranas del Alzheimer.

Como ya se ha mencionado, estos datos serán particionados en dos subgrupos: *training* y *test*. Para el caso de la validación *hold-out*, un 75% de los datos serán empleados para el *training*, lo que equivale a un total de 113 sujetos, mientras que los datos de los 37 sujetos restantes serán reservados para la fase de prueba. Por otra parte, para la validación cruzada se emplearán la totalidad de los datos tanto para la fase de entrenamiento como para la de prueba. La metodología para llevar a cabo la validación cruzada será explicada en mayor detalle en el apartado de *Métodos*. Las edades de estos sujetos están comprendidas entre 57,2 y 91,4 años. La media de la edad es de 74 años con una desviación típica de $\pm 7,97$ años. Los datos fueron descargados de la base de datos de la Iniciativa de neuroimagen de la enfermedad de Alzheimer, comúnmente conocida como ADNI por sus siglas en inglés (Petersen *et al.*, 2009)

3.2 Software ALTEA

Para el desarrollo de la herramienta de clasificación del software ALTEA (López Gómez, 2017) se empleó el entorno de programación MATLAB®, concretamente la versión R2015b. A su vez, la interfaz gráfica de usuario se creó mediante la aplicación GUIDE, ofrecida por el propio MATLAB.

Además, para implementar las funciones de los modelos de clasificación se emplearon librerías de MATLAB, concretamente fueron basadas en el código exportable de la aplicación de MATLAB *Classification Learner App*. A su vez, con el fin de complementar dicho código y gracias a la comunidad de usuarios de MATLAB, se utilizó una librería en la que se encontraban implementados métodos de selección de características como el ReliefF o Fisher, llamada *FSLib 2017*, elaborada por el usuario Giorgio Roffo de la comunidad de MATLAB (Roffo, 2017; Roffo *et al.*, 2017; Roffo y Melzi, 2017).

4. MÉTODOS

Recordemos que, el principal objetivo de este trabajo reside en el análisis de nuevos biomarcadores extraídos a partir de texturas de IRM de pacientes control (CN), pacientes con deterioro cognitivo leve temprano (EMCI) y pacientes con enfermedad de Alzheimer (AD). Para ello, se pretende desarrollar una herramienta que permita evaluar características extraídas de dichas imágenes empleando distintas técnicas de aprendizaje máquina con el fin de encontrar biomarcadores que puedan ser significativos en la detección de Alzheimer en fases tempranas. Además, empleando esta herramienta y a partir de dichos subconjuntos de datos, se entrenarán y testearán distintos modelos de evaluación de características para evaluar la capacidad predictiva de dichas texturas.

A continuación, se explicará la metodología seguida para la evaluación de nuevos biomarcadores. Partiendo de un *data set* conformado por 150 imágenes de pacientes CN, AD y EMCI esta se puede resumir en las siguientes actuaciones:

- I. Segmentación de la ROI (hipocampo) en 2D y 3D
- II. Extracción y análisis de texturas empleando histograma, GLCM y GRLM
- III. Visualización de los datos
- IV. Clasificación

Este trabajo se centrará especialmente en la evaluación de las texturas extraídas, para ello, se han desarrollado dos módulos: uno empleado para la visualización de datos y otro para su clasificación.

4.1 Segmentación y análisis de texturas

Previamente al análisis de las texturas se requiere de la extracción de las mismas a partir de la ROI (hipocampo) de imágenes de resonancia magnética cerebral 2D y 3D de pacientes control, pacientes con deterioro cognitivo leve temprano y pacientes con enfermedad de Alzheimer. El análisis de texturas se realizará sobre cada una de las seis regiones de interés previamente obtenidas, que serán referidas como ROI 1_D, ROI 2_D y ROI 3_D para el hipocampo derecho, y ROI 1_I, ROI 2_I y ROI 3_I para el izquierdo.

La segmentación de dichas ROI se realizó empleando regiones circulares o esféricas. El radio de la ROI para ROI 1_D y ROI 1_I fue de 3 píxeles, por otra parte, para ROI 2_D y ROI 2_I se seleccionó un radio de 5 píxeles y, por último, la segmentación de ROI 3_D y ROI 3_I se realizó con círculos o esferas de 8 píxeles de radio.

A partir de dicha segmentación, se extrajeron una serie de texturas empleando el paquete *RADIOMICS* disponible en la librería de MATLAB, implementado por Vallières et al (Vallières *et al.*, 2015). El término *radiomics* hace referencia al campo de estudio de datos extraídos a partir de imágenes médicas. Concretamente, se obtuvieron un total de 25 características (Tabla 3) para el hipocampo segmentado de cada paciente, empleando los métodos del histograma de primer orden, la matriz de largo recorrido (GLRLM) y la matriz de co-ocurrencia (GLCM). Las ecuaciones para extraer dichas características se encuentran recogidas en el artículo anteriormente mencionado (Vallières *et al.*, 2015).

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Método de análisis de texturas	Escala	Número de características	Nombre de las características
Histograma	Global	3	<i>Variance, Skewness, Kurtosis</i>
Matriz de co-ocurrencia (GLCM)	Local	9	<i>Energy, Contrast, Entropy, Homogeneity, Correlation, SumAverage, Variance, Dissimilarity, AutoCorrelation</i>
Matriz de largo recorrido (GLRLM)	Regional	13	<i>Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray-level Non-uniformity (GLN), Run-Length Non-uniformity (RLN), Run Percentage (RP), Low Gray-level Run Emphasis (LGRE), High Gray-level Run Emphasis (HGRE), Short Run Low Gray-level Emphasis (SRLGE), Short Run High Gray-level Emphasis (SRHGE), Long Run Low Gray-level Emphasis (LRLGE), Long Run High Gray-level Emphasis (LRHGE), Gray-level Variance (GLV), Run-Length Variance (RLV)</i>

Tabla 3. Texturas obtenidas para cada ROI

Estos parámetros se encuentran recogidos en tablas nombradas como ROI 1_D_2D, ROI 1_I_2D, ROI 2_D_2D, ROI 2_I_2D, ROI 3_D_2D, ROI 3_I_2D, ROI 1_D_3D, ROI 1_I_3D, ROI 2_D_3D, ROI 2_I_3D, ROI 3_D_3D, ROI 3_I_3D que son importadas a la herramienta de evaluación de características con el fin de analizar la presencia de nuevos biomarcadores para la detección precoz de Alzheimer.

4.2 Visualización de datos

El primer módulo que encontramos dentro de la herramienta de evaluación de características consiste en un diagrama de dispersión o *scatter plot* que se emplea para la visualización de los datos en función de las variables escogidas (Figura 11).

Las gráficas de dispersión son un tipo de diagrama matemático de puntos que representa la relación entre dos variables para un conjunto de datos. El principal objetivo de la presencia de esta gráfica observar cómo están distribuidas espacialmente las muestras en función de las texturas escogidas, lo que permite tener indicios acerca de cómo serán los resultados de la posterior clasificación.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

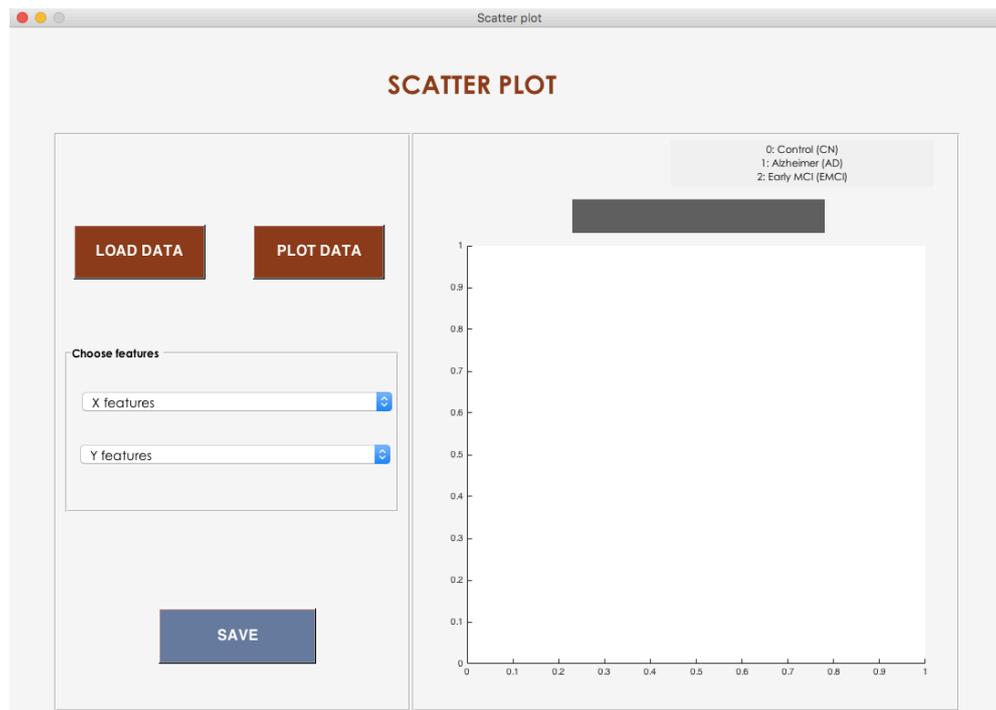


Figura 11. Pantalla inicial gráfica de dispersión.

La metodología de uso de esta herramienta consiste en que, en primer lugar, pulsando el botón *LOAD DATA* se carga en la base de datos una tabla en formato *.xlsx* o *.xls* (Figura 12) que contiene distintas texturas extraídas a partir de las IMR de distintos tipos de pacientes (CN, AD y EMCI). Una vez estos datos han sido cargados en la herramienta, en el panel izquierdo de la aplicación se escogen cuáles son las variables en función de las cuales queremos representar los datos. A continuación, se grafica el resultado siguiendo un diagrama de dispersión, tal y como se muestra en la Figura 13.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

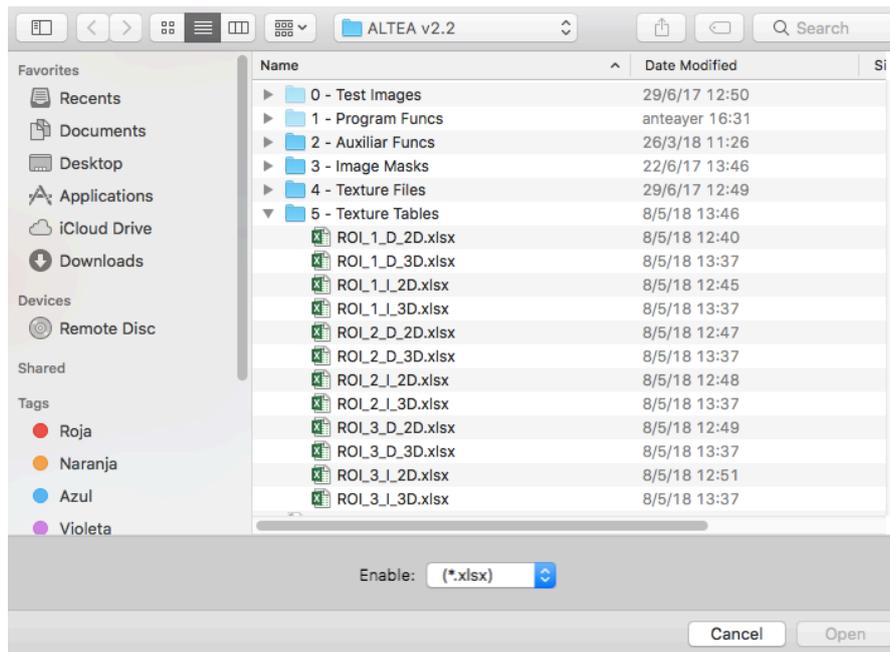


Figura 12. Carga de datos en la herramienta.

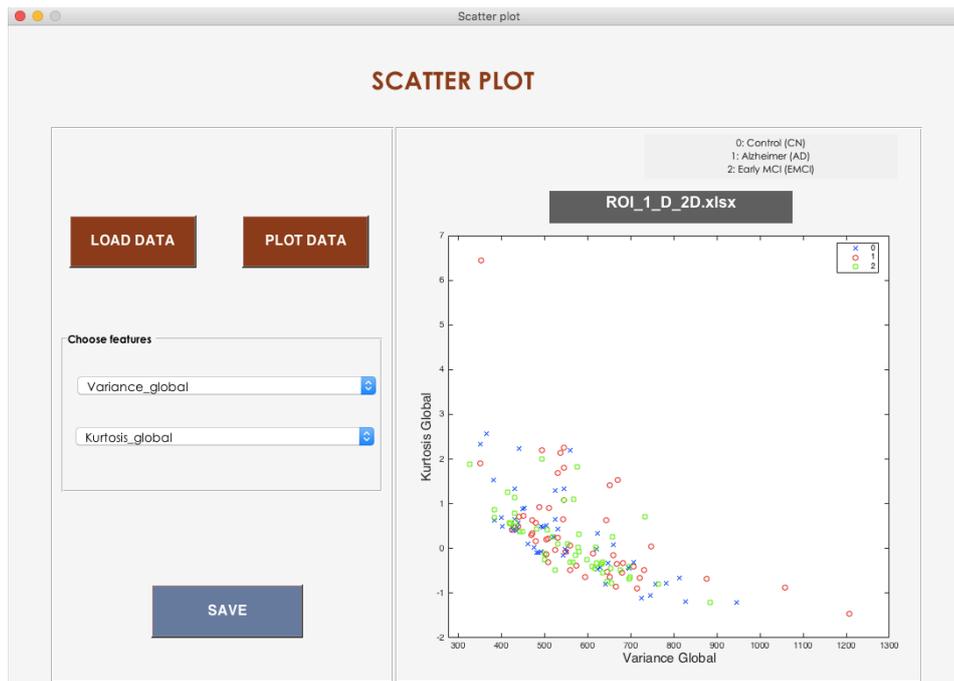


Figura 13. Visualización de ejemplo de gráfica de dispersión en función de la varianza y la curtosis para los datos de la tabla ROI_1_D_2D.

Como se puede observar en la Figura 13, los pacientes control están representados con aspas azules, los que tienen enfermedad de Alzheimer con círculos rojos y los de deterioro cognitivo leve temprano con círculos verdes.

4.3 Clasificación

Durante esta sección, se ha elaborado una herramienta de clasificación basada en técnicas de aprendizaje automático. Esta herramienta, cuya interfaz gráfica se observa en la Figura 14, elabora un ranking con los parámetros que resultan más significativos en la discriminación de los datos en distintos grupos (AD, CN y EMCI) y, a partir de dichos parámetros, elabora un modelo clasificador que permita un diagnóstico del Alzheimer temprano y preciso.

El clasificador tiene como principal objetivo determinar y comparar el valor predictivo de los distintos modelos generados. Para ello, permite comparar indicadores como el porcentaje de exactitud (*percentage of accuracy*), la curva ROC, el área bajo la curva ROC (AUC) y la matriz de confusión. De este modo, a partir de la evaluación y comparación de dichos indicadores, se determina cuál es el mejor modelo automático predictivo.

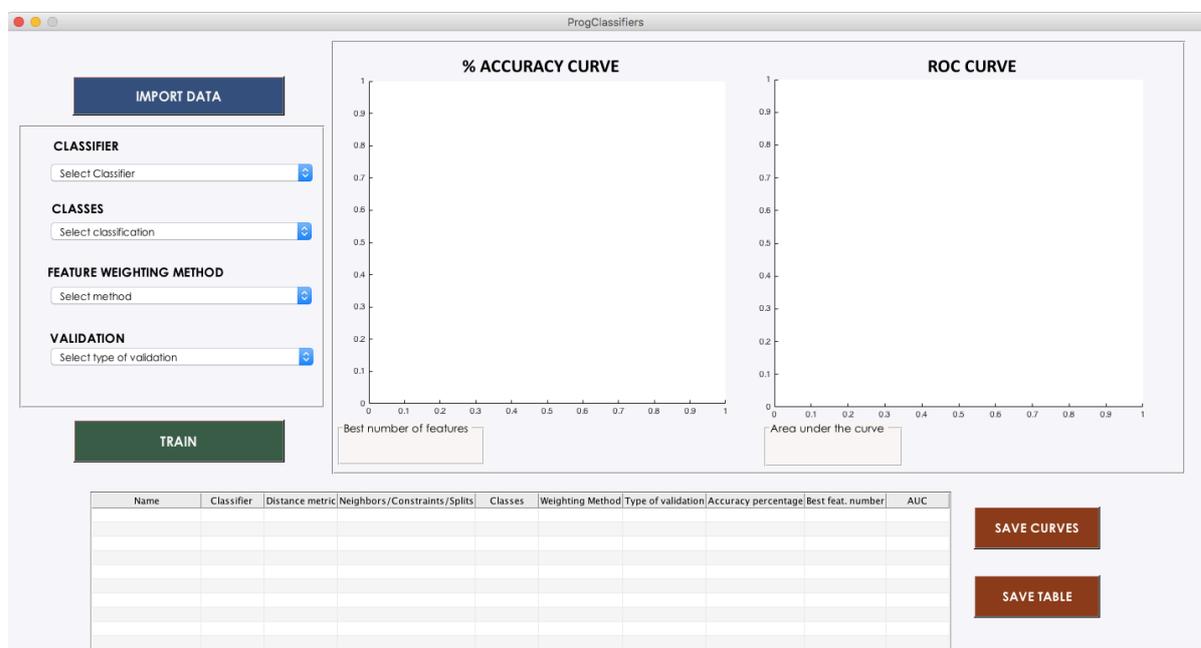


Figura 14. Pantalla inicial del clasificador.

La metodología de uso de esta herramienta consiste en que, al igual que en el módulo anterior, en primer lugar, se debe importar una tabla de datos en formato *.xls* o *.xlsx* que recoja, para una misma ROI, todas las texturas para pacientes de las tres clases (ver Figura 12). Una vez cargada, el modelo se elaborará escogiendo un método de clasificación, los grupos a clasificar, un método de selección de características y un método de validación:

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

- *Classifier* o método de clasificación. Existen tres opciones de métodos clasificatorios a implementar:

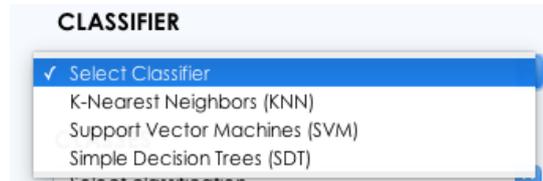


Figura 15. Pantalla de opción de clasificador.

- *K-Nearest Neighbors (KNN)*. Recordemos que, este método se basa en calcular la distancia de la nueva muestra a una serie de vecinos (K), para ello, se puede calcular la distancia de varias formas posibles, a la vez que se pueden ponderar los pesos con los que influyen cada vecino a la nueva muestra. Al seleccionar este método dentro de la herramienta, surgirán una serie de pantallas emergentes que solicitarán al usuario distintas entradas (Figuras 16, 17 y 18):
 - *Number of neighbors*. Este término se refiere al número de vecinos más cercanos con los que comparar la nueva muestra.

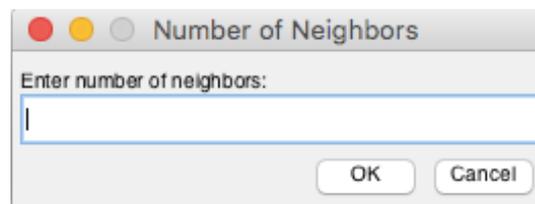


Figura 16. Pantalla de la opción KNN que solicita el número de vecinos.

- *Distance metric*. Hace referencia a la fórmula con la que calcular la distancia entre muestras, a elegir entre: distancia euclídea, distancia de *Chebyshev* y distancia de *Manhattan* o *Cityblock* (ver fórmulas en el apartado 2.4.1.1)

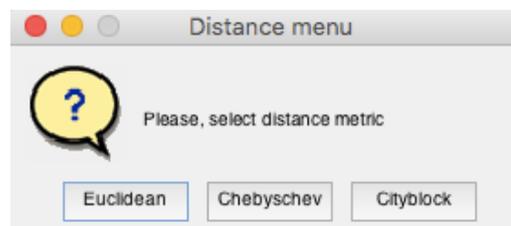


Figura 17. Pantalla de la opción KNN que solicita la métrica para calcular la distancia.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

- *Support Vector Machines (SVM)*. Como ya se comentó, este método define un hiperplano a modo de frontera para clasificar los datos en clases que debe estar lo más separado posible de ambas clases (margen). Además, para poder definir dicha frontera en casos no lineales se requiere de una función denominada *kernel*. Para utilizar el SVM como clasificador, será necesario introducir una serie de especificaciones:
 - *Kernel function*. Este concepto especifica cuál es el tipo de función *kernel* que usará nuestro clasificador SVM. En concreto, se puede escoger entre: *kernel* lineal (*Linear*), Gaussiano (*Gaussian or Radial Basics Function (RBF)*) y cuadrático (*Quadratic*).



Figura 18. Pantalla de la opción SVM que solicita la función kernel deseada para el modelo.

- *Box constraint level*. Este parámetro se encuentra estrechamente relacionado con el parámetro de costo de los SVM. Se emplea para restringir los valores posibles para los multiplicadores de Lagrange, de este modo, al aumentar el nivel de restricción (*box constraint level*) se logra conseguir una disminución de los vectores de soporte, aunque normalmente aumenta el tiempo de entrenamiento.

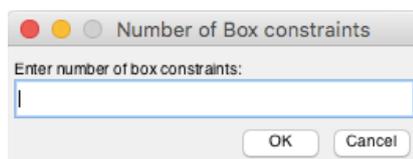


Figura 19. Pantalla de la opción SVM que solicita el número de restricciones.

Cabe destacar que, la clasificación SVM fue desarrollada inicialmente para resolver únicamente problemas de clasificación binarios. Sin embargo, también es posible resolver problemas multiclase, gracias a que se puede reducir el problema de clasificación multiclase a un conjunto de subproblemas de clasificación binarios, con un modelo de SVM para cada subproblema. En estos casos, tal y como se ha comentado previamente, la clasificación puede realizarse *one-vs.-one*, donde se entrena a un modelo por cada par de clases (aprende a distinguir una clase de otra), o *one-vs.-all*, donde se entrena a un modelo para cada clase (aprende a distinguir una clase

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

de todas las demás). En la opción de SVM de esta herramienta, se ha escogido por defecto una clasificación *one-vs.-all*.

- *Simple Decision Trees (DT)*. Se trata de la clasificación más rápida y fácil de interpretar, aunque puede conllevar problemas con la precisión del modelo. En la herramienta, para escoger esta opción se solicitará que se introduzca el número máximo de divisiones o puntos de ramificación para elaborar el árbol de decisión y así controlar su profundidad (Figura 20). A la hora de llevar a cabo este tipo de clasificación, se debe considerar que los árboles de decisión son métodos simples, que ocupan poca memoria, rápidos y sencillos de interpretar, sin embargo, se debe tener siempre en cuenta que, dependiendo de factores como su diseño, el tamaño muestral y la calidad de los datos, pueden conllevar a predicciones poco precisas.

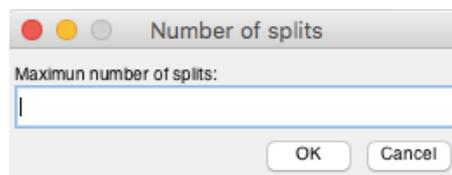


Figura 20. Pantalla de la opción DT que solicita el número máximo de divisiones.

Un árbol con muchas hojas y muy refinado suele ser muy preciso con sus datos de entrenamiento, sin embargo, no siempre se cumple que exista este grado de precisión en la clasificación de conjunto de prueba independiente. Esto es consecuencia del sobreentrenamiento que suele aparecer en árboles de decisión con estas características.

Por el contrario, un árbol más simple, que no alcance una alta precisión de entrenamiento puede llegar a ser más robusto, pues su porcentaje de *accuracy* en los datos de entrenamiento puede asimilarse más al de los datos del conjunto de prueba. Además, será más fácil de interpretar.

- *Classes*. Es necesario establecer si se trata de un problema de clasificación binario o multiclase. Ambas opciones se encuentran disponibles en la herramienta. Se puede realizar una clasificación para las clases Alzheimer vs. Control (AD-CN), Control vs. Deterioro Cognitivo Leve Temprano (CN-EMCI), Alzheimer vs. Deterioro Cognitivo Leve Temprano (AD-EMCI) o Multiclase (*multiclass*).

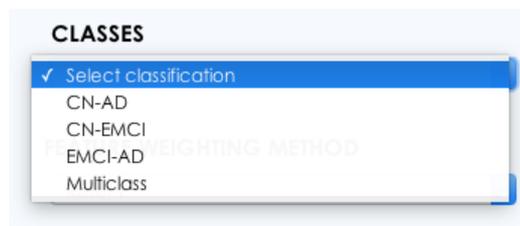


Figura 21. Opción de selección de las clases del problema.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

- *Feature weighting method* o método de ponderación de características. A la hora de elaborar un ranking mediante *feature selection* que determine cuáles son las características más discriminativas para la clasificación, se pueden escoger tres métodos distintos, vistos en la sección 2.4.2:
 - Relief-F.
 - Fisher.
 - P-value.



Figura 22. Opción de selección del método de ponderación de características.

- *Validation* o tipo de validación. La herramienta permite escoger entre realizar una validación de tipo *hold-out* o una validación cruzada (*cross validation*).

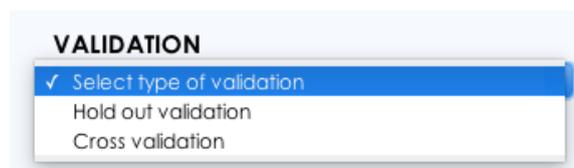


Figura 23. Opción de selección del método de validación.

Como ya se ha comentado, en ambos tipos de validación se realiza una partición del total de datos en subgrupos de datos de training y test. En el caso de la validación *hold-out* de esta herramienta, a la hora de obtener los datos de test se escoge por defecto un porcentaje del 25% de la muestra total, de modo que, por ejemplo, para el caso *multiclass* si se tiene un total de 150 muestras, 113 irán destinadas a entrenar el modelo y 37 se emplearán como datos de prueba, mientras que para casos binarios en los que únicamente se cuenta con 100 muestras, 75 serán de *training* y las 25 restantes de *test*. El modelo será entrenado con los predictores que el ranking haya considerado más exitosos. A partir de dicho entrenamiento, se obtiene el vector de porcentajes de *accuracy* (representado en la gráfica de la izquierda para cada característica), a continuación, se calcula el máximo valor de dicho vector y se reentrena el modelo con las características que se corresponden para dicho valor máximo de *accuracy*.

En el caso de la implementación de la validación cruzada en esta herramienta, se emplea el método *k-fold* para llevar a cabo 5 particiones del total de datos. A continuación, se elabora un ranking para cada uno de esos cinco subgrupos y se entrena el modelo. Como resultado, tendremos cinco vectores con el porcentaje de *accuracy*. Para obtener un único vector, se hace la media de estos y, posteriormente, se seguirá el procedimiento habitual, similar al del método *hold-out*: se calcula el máximo del vector de porcentajes de exactitud, se obtienen los predictores correspondientes para dicho valor y se reentrena el modelo con ellos.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Con el objetivo de que se encuentre registrado todo entrenamiento realizado con la herramienta y sus respectivos resultados, en la mitad inferior de la interfaz gráfica se encuentra una tabla que recoge dicha información (ver Figura 24).

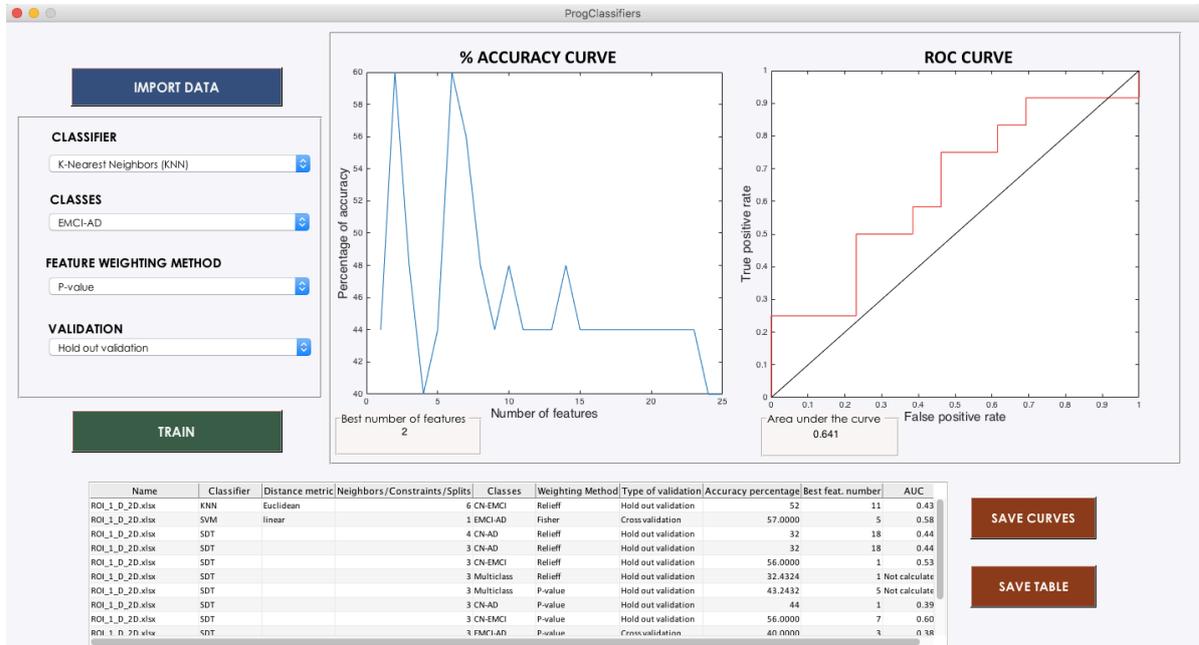


Figura 24. Clasificador en funcionamiento.

Para evaluar los resultados de la clasificación del modelo seleccionado, se estudiarán principalmente los parámetros del porcentaje de *accuracy* y el área debajo de la curva ROC (AUC), ambos recogidos en la tabla. Además de la tabla, la herramienta cuenta con un panel superior en el que se representan dos gráficas, una correspondiente con los valores de porcentajes de *accuracy* para cada subconjunto de características y otra correspondiente con la curva ROC, tal y como se observa en la Figura 24. En la gráfica ROC aparece una línea en negro a modo de referencia para facilitar la interpretación de la curva. Cabe recordar que, tal y como se comentó previamente, la curva ROC se define principalmente para clasificaciones binarias (aunque actualmente existen estudios que tratan de desarrollarla para problemas de más de dos clases); es por ello que, en el caso de la clasificación multiclase la gráfica de esta curva no aparece y, por lo tanto, tampoco se puede calcular el valor de AUC. Para estudiar la clasificación de problemas multiclase únicamente se evaluará el porcentaje de *accuracy* y la matriz de confusión.

Cabe destacar que, gracias al botón *SAVE TABLE* de la interfaz gráfica, estas tablas pueden ser exportadas para ser visualizadas en programas como Microsoft Excel® para una mayor comodidad durante su análisis y visualización.

Además, como ya se ha comentado, para evaluar el ranking realizado con el *feature weighting method* seleccionado y con el fin de liberar al programa de carga, se realiza un reentrenamiento del programa con las características que mejor clasificación ofrecen observando los valores de *accuracy*. En la gráfica izquierda del panel superior se observa la curva *profile* la cual

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

se obtiene durante el proceso de *feature selection* y permite identificar el subconjunto de características que permiten la mejor clasificación.

Con el principal objetivo de estudiar los resultados de la clasificación, además de contar con la tabla y las curvas del panel superior de la herramienta, cada vez que se lleve a cabo una nueva clasificación, surgirá una nueva figura que contenga una matriz de confusión (ver Figura 25). Las celdas diagonales en color verde de la matriz se corresponden con las observaciones correctamente clasificadas, mientras que las celdas fuera de dicha diagonal (en rojo) representan observaciones incorrectamente clasificadas, además, en cada celda se observa tanto el número de muestras clasificadas como el porcentaje que dichas muestras representan con respecto al total. La celda en la parte inferior derecha de la gráfica (en color azul) muestra el porcentaje de *accuracy* de la clasificación general.

En cuanto a la columna que aparece en el extremo derecho en color gris, se muestran los porcentajes de todos los ejemplos que se predice que pertenecen a cada clase que están clasificados correcta e incorrectamente. Estas métricas son conocidas como valor predictivo positivo, en color verde, y tasa de descubrimiento falso, en rojo. La fila gris en la parte inferior de la gráfica muestra los porcentajes de todos los ejemplos que pertenecen a cada clase que están clasificados correcta e incorrectamente (sensibilidad o tasa positiva verdadera y especificidad o tasa de falsos negativos). Estas métricas fueron vistas en el punto 2.4.4.1, donde se pueden consultar las fórmulas empleadas para su cálculo.

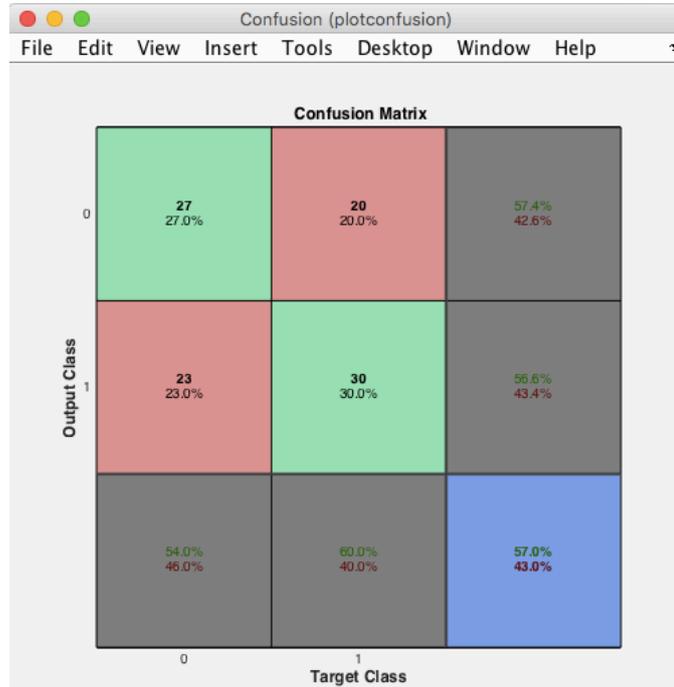


Figura 25. Ejemplo matriz de confusión.

5. RESULTADOS Y DISCUSIÓN

5.1 Desarrollo de herramienta para la visualización de datos

El primer resultado de este trabajo reside en el desarrollo de una interfaz para la visualización de datos que puede encontrarse dentro del programa ALTEA (Alzheimer TExture Analyzer) el cual se comenzó a desarrollar durante un Trabajo de Fin de Grado del pasado curso 2017/2018 llamado *Análisis y evaluación de nuevos biomarcadores para la enfermedad de Alzheimer mediante un proceso radiomics basado en el análisis de texturas sobre imágenes de resonancia magnética* (López Gómez, 2017). La herramienta *scatter plot* permite estudiar la correlación entre texturas de IMR mediante un gráfico de dispersión. En primer lugar, el usuario importa tablas donde se encuentran recogidas una serie de texturas previamente extraídas y posteriormente escoge cuál es la comparación que desea realizar. El gráfico de dispersión que se origina puede ser guardado en todo momento como una imagen mediante el botón *SAVE*.

Este análisis se realizará para las 12 regiones de interés (ROI), de las cuales 6 serán circulares (imágenes 2D) y 6 esféricas (imágenes 3D), cada una de ellas con un radio de 3, 5 y 8 píxeles respectivamente, tanto para la región izquierda del hipocampo como para la derecha. Como ya se ha comentado, el análisis se realizará en una muestra con un total de 150 pacientes (50 del grupo CN, 50 del grupo AD y 50 del grupo EMCI).

En este caso, la herramienta *scatter plot* actúa como herramienta de pre-visualización de los datos, de modo que permite al usuario apreciar si existe una separación visible entre grupos de pacientes control (CN), pacientes con deterioro cognitivo (EMCI) y pacientes con enfermedad de Alzheimer (AD). Una buena separación de grupos podría ser indicativa de que la posterior clasificación con *Classifiers* será de buena calidad.

5.1.1 Visualización texturas 2D

Tras estudiar la tendencia general de la expresión de los datos en las 6 tablas de texturas 2D, se determinó que al graficar la distribución de datos empleando las texturas de *kurtosis* y *variance* globales, los datos mostraban una ligera tendencia de agrupamiento. Esta tendencia se apreció en todas las tablas de texturas 2D, aunque se hizo más notable en el caso de ROI_2_I_2D.xlsx (Figura 26).

En este caso en concreto se observa como las muestras de los pacientes con AD (círculos rojos) se encuentran algo más separadas de las muestras CN (aspas azules) y EMCI (círculos verdes). En cambio, las muestras EMCI y CN se encuentran distribuido en el mismo espacio 2D basado en estas dos características. Este resultado muestra indicios de que la clasificación entre muestras CN y EMCI usando estas 2 características no será muy satisfactoria y en cambio la clasificación de muestras CN o EMCI vs. AD podría ser mejor.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

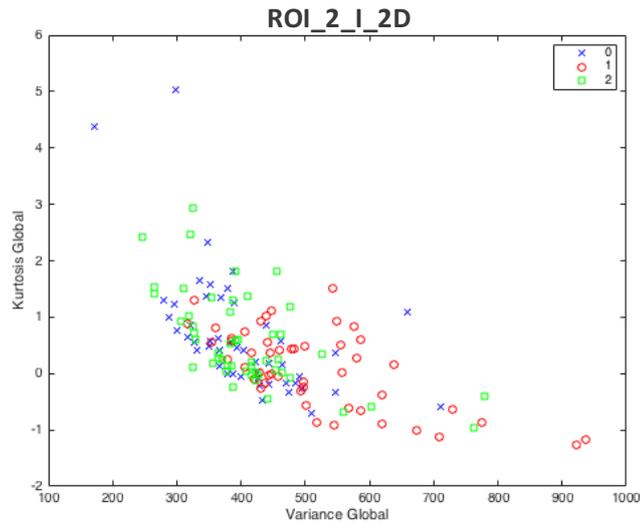


Figura 26. Diagrama de dispersión de *Variance_global* y *Kurtosis_Global* para ROI_2_I_2D. En aspas azules se representan los pacientes CN (0), con círculos rojos los pacientes AD (1) y con círculos verdes los EMCI (2).

Por otro lado, durante el estudio de la representación de los distintos datos en el espacio característico formado por las variables de varianza y curtosis globales también se observaron gráficas donde no se apreciaba una diferenciación clara entre grupos de muestras (Figura 27) y, por lo tanto, indicaron que la posterior clasificación con no sería del todo buena.

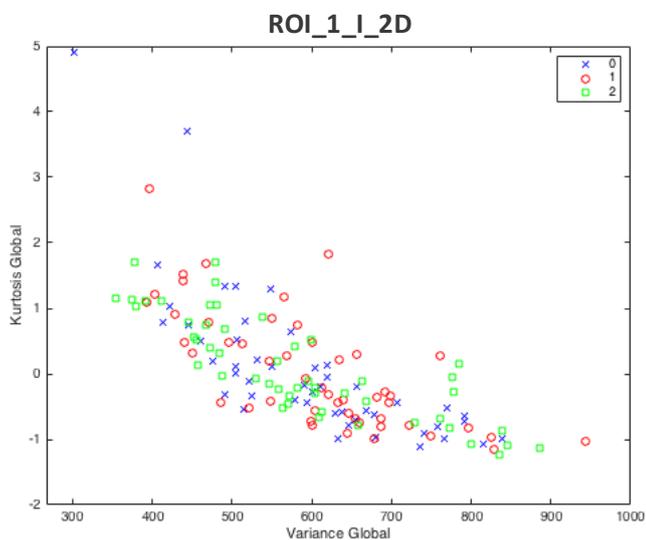


Figura 27. Diagrama de dispersión de *Variance_global* y *Kurtosis_Global* para ROI_1_I_2D. En aspas azules se representan los pacientes CN (0), con círculos rojos los pacientes AD (1) y con círculos verdes los EMCI (2).

5.1.2 Visualización texturas 3D

En cuanto al estudio de la expresión de los datos 3D, se volvió a determinar que estos mostraban una ligera tendencia de agrupamiento al ser representados en un espacio característico conformado por las variables de *kurtosis* y *variance* globales.

Tras analizar las distintas gráficas de dispersión se volvió a apreciar como la mejor clasificación se prevé para los grupos AD vs. CN o AD vs. EMCI. Un ejemplo de ello se aprecia en la Figura 28, la cual muestra la distribución de los datos para una región de interés de 5 píxeles. A su vez, también se volvieron a encontrar gráficas que no mostraban una diferenciación clara entre clases (Figura 29). Cabe destacar que, al igual que ocurría en el caso de las texturas 2D, las gráficas que mostraron una peor clasificación fueron aquellas obtenidas con la ROI de menor radio (3 píxeles).

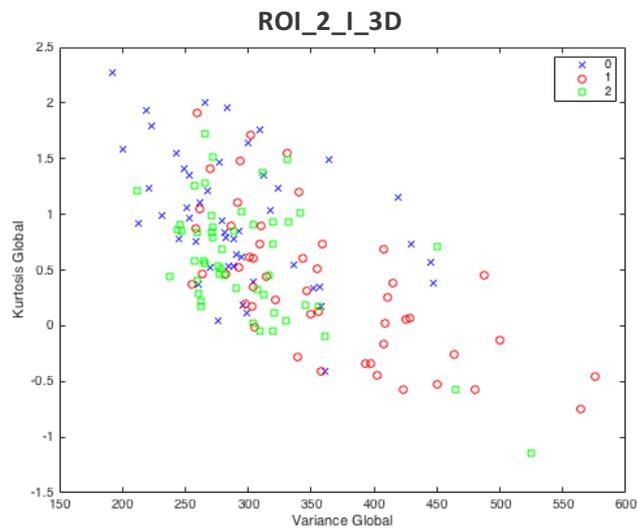


Figura 28. Diagrama de dispersión de *Variance_global* y *Kurtosis_Global* para ROI_2_I_3D.

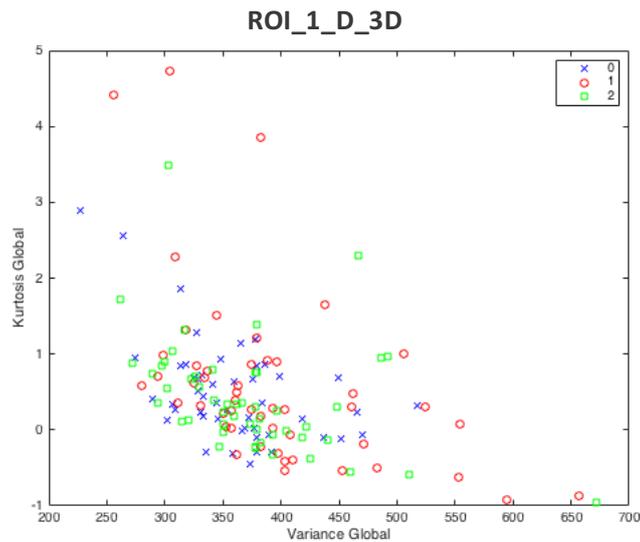


Figura 29. Diagrama de dispersión de *Variance_global* y *Kurtosis_Global* para ROI_1_D_3D.

5.2 Desarrollo *Classifiers*

El segundo y principal resultado de este trabajo consiste en la implementación de una herramienta clasificadora que permite analizar las texturas de imágenes cerebrales adquiridas mediante RM mediante técnicas de aprendizaje máquina, bautizada como *Classifiers*. Esta herramienta termina de conformar el programa ALTEA, aportando un método preciso de evaluación de características, su interfaz gráfica fue mostrada durante el apartado de *Métodos*, concretamente en la Figura 14 se aprecia la pantalla inicial del clasificador y en la Figura 24 la interfaz del clasificador en funcionamiento.

Gracias a *Classifiers*, se logrará llevar a cabo una distinción de clases a partir del entrenamiento previo del programa con distintos datos de texturas, lo que permitirá concluir cuál es el modelo de clasificación más apto para diferenciar cada clase, así como si existe alguna mejora en la clasificación en función del tamaño de la región de interés o si se emplean regiones en dos o tres dimensiones.

5.3 Análisis de texturas

El objetivo principal consiste en hallar nuevos biomarcadores que permitan la detección de la enfermedad de Alzheimer empleando la herramienta *Classifiers*. Para analizar las texturas en imágenes 2D y 3D, comenzaremos realizando una clasificación multiclase en la que se analizarán grupos control (CN), Alzheimer (AD) y deterioro cognitivo leve (EMCI). Una vez analizada la clasificación multiclase, se realizará un estudio sobre la clasificación binaria de estos mismos grupos: CN vs. AD, EMCI vs. AD y CN vs. EMCI.

Se trata de buscar el modelo que mejor realice la clasificación de los datos de las 12 ROI's en cada grupo: CN, AD o EMCI. Para determinarlo, se evaluará la tabla generada, las curvas que aparecen en la interfaz y la matriz de confusión que surge como nueva figura en cada clasificación.

Seleccionando un método de clasificación, las clases de estudio, un método de ponderación de características y un método de validación podremos elaborar un modelo. Cabe destacar que se trata de un análisis complejo debido al alto número de combinaciones posibles de las que se dispone. Es por ello que, para evaluar los resultados se harán pruebas para cada región introduciendo determinados parámetros para elaborar el modelo clasificador. El número de vecinos (para KNN), *box constraints* (para SVM) o *splits* (para SDT) utilizado para realizar este proyecto se recoge en la siguiente tabla:

KNN	Nº de vecinos	3	7	25	
SVM	Nº <i>box constraints</i>	1	100		
SDT	Nº <i>splits</i>	3	5	10	25

Tabla 4. Parámetros para elaborar los modelos analizados

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

En cuanto a la muestra de prueba y de entrenamiento empleada, un 25% de los sujetos se emplearán como datos de prueba en el caso de la validación *hold-out*. Sin embargo, a pesar de que la herramienta *Classifiers* dispone de la opción *hold-out*, los resultados de este trabajo se centrarán en el tipo de clasificación *cross-validation*, debido a que, como ya se comentó durante el punto de *Introducción*, la técnica de validación *hold-out* conlleva ciertas desventajas, entre ellas está la toma de datos aleatoria para elaborar cada conjunto; es por ello que, el porcentaje de *accuracy* calculado con esta técnica no es del todo fiable.

5.3.1 Análisis de texturas 2D

Los resultados del análisis de texturas en dos dimensiones obtenidos a partir de la elaboración de modelos combinando las posibles opciones y parámetros vistos en la Tabla 4 fueron exportados a distintos documentos de excel. Para determinar cuál fue el mejor modelo para cada una de las 6 regiones de interés 2D se valoró el porcentaje de exactitud: aquellos modelos con el valor de *accuracy* más alto fueron seleccionados como los mejores. En caso de existir varios modelos con un mismo valor para el porcentaje de *accuracy*, prevaleció aquel que lograra dicho porcentaje para un menor número de características. Si, tras aplicar estos criterios, continúan existiendo dudas, se tendrá en cuenta las diferencias existentes para el valor de AUC, de forma que será seleccionado el modelo con mayor AUC. Todos los resultados mostrados a continuación fueron obtenidos a partir del método de validación cruzada por razones que ya se han comentado.

Cabe destacar que, para estudiar la clasificación multiclase se empleará principalmente el porcentaje de exactitud y la matriz de confusión, pues no se calculará la curva ROC ni el área debajo de esta, razón por la cual no aparece el valor de AUC en la Tabla 5.

Además, al tratarse de la evaluación de grandes cantidades de resultados y considerando la extensión limitada de este proyecto, se tratará de analizar principalmente aquellos que hayan destacado dentro de su respectivo conjunto. Estos serán analizados desde el punto de vista de la magnitud del porcentaje de *accuracy*, la mejor clasificación según las clases comparadas y la mejor clasificación para cada región.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Clases	Región de interés	Método de clasificación	Tipo de distancia	Nº vecinos, <i>Box const.</i> o <i>splits</i>	Método de ponderación	Accuracy (%)	Nº de características	AUC
Multiclass	ROI_1_D_2D	SVM	Linear	1	ReliefF	43,33	21	-
	ROI_1_I_2D	SVM	Gaussian	1	Fisher	43,33	6	-
	ROI_2_D_2D	SVM	Gaussian	1	P-value	54,00	24	-
	ROI_2_I_2D	SVM	Linear	1	ReliefF	52,00	21	-
	ROI_3_D_2D	SVM	Linear	1	ReliefF	46,00	19	-
	ROI_3_I_2D	SVM	Linear	100	P-value	51,33	15	-
CN vs. AD	ROI_1_D_2D	SVM	Linear	1	ReliefF	65,00	18	0.68
	ROI_1_I_2D	SVM	Linear	1	P-value	65,00	7	0.65
	ROI_2_D_2D	SVM	Linear	1	Fisher	72,00	25	0.78
	ROI_2_I_2D	KNN	Euclidean	3	P-value	76,00	16	0.82
	ROI_3_D_2D	SVM	Linear	1	ReliefF	66,00	21	0.72
	ROI_3_I_2D	KNN	Chebychev	7	ReliefF	67,00	12	0.72
CN vs. EMCI	ROI_1_D_2D	SVM	Linear	1	Fisher	63,00	11	0.62
	ROI_1_I_2D	KNN	Euclidean	3	P-value	66,00	1	0.63
	ROI_2_D_2D	SVM	Gaussian	1	P-value	65,00	3	0.67
	ROI_2_I_2D	KNN	Euclidean	3	P-value	62,00	1	0.62
	ROI_3_D_2D	KNN	Euclidean	3	P-value	57,00	5	0.62
	ROI_3_I_2D	KNN	Euclidean	7	ReliefF	64,00	15	0.67
EMCI vs. AD	ROI_1_D_2D	SVM	Linear	100	P-value	60,00	11	0.63
	ROI_1_I_2D	SVM	Gaussian	1	P-value	62,00	3	0.61
	ROI_2_D_2D	SDT	-	3	P-value	73,00	5	0.76
	ROI_2_I_2D	KNN	Chebychev	7	P-value	78,00	16	0.77
	ROI_3_D_2D	SVM	Gaussian	1	P-value	71,00	20	0.73
	ROI_3_I_2D	KNN	Euclidean	3	P-value	71,00	9	0.69

Tabla 5. Resultados para regiones de interés de 2 dimensiones.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Como el principal factor que tendremos en cuenta para evaluar la clasificación del modelo será el porcentaje de exactitud de este, para una mejor visualización de los resultados, se elaboró una gráfica con los porcentajes de *accuracy* para cada clasificación (Figura 30).

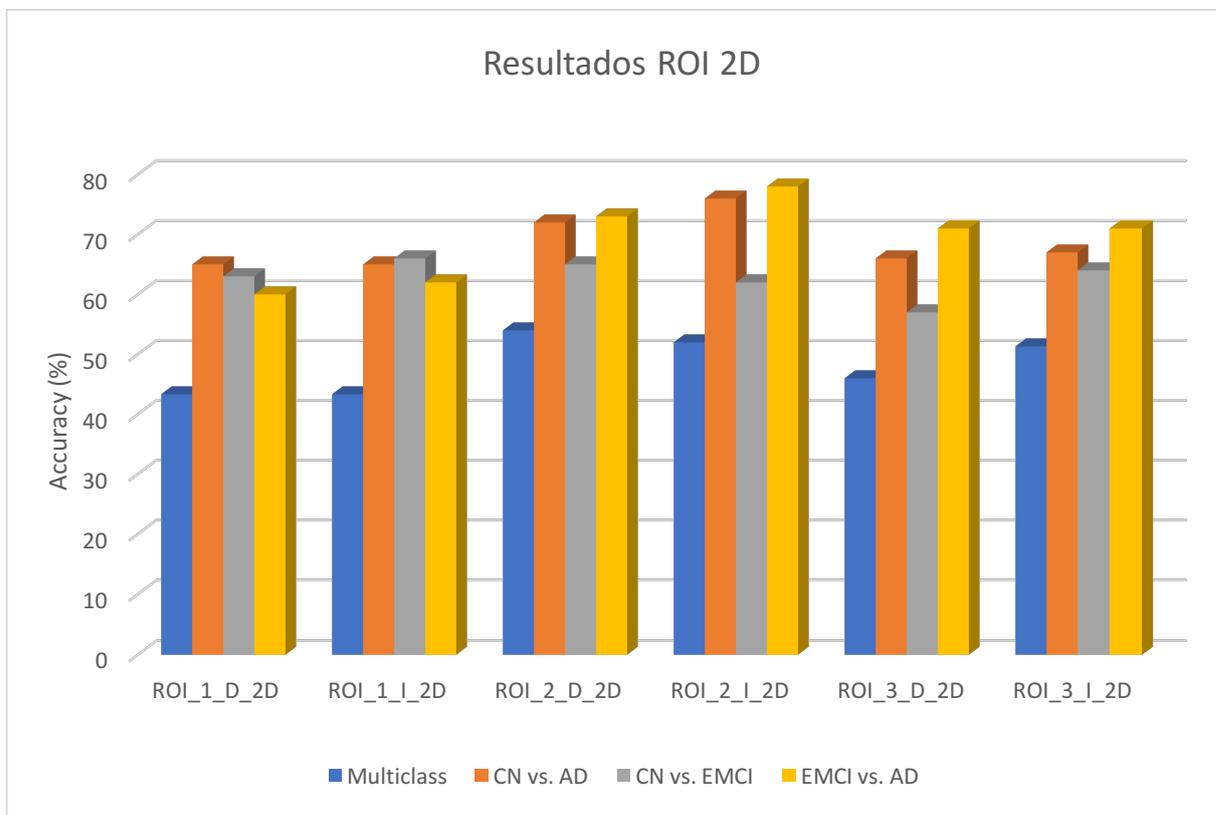


Figura 30. Visualización resultados de regiones de interés de 2 dimensiones.

Estudiando detenidamente los resultados se observa como los valores de exactitud más altos se alcanzan para la clasificación EMCI vs. AD de la ROI_2_I_2D, obtenida con 16 de las 25 características correctamente ordenadas según su *p*-valor. Tal y como se aprecia en la Figura 31, para una clasificación binaria, *Classifiers* llegó a predecir con una precisión del 71% 36 casos de pacientes con deterioro cognitivo leve temprano y 35 pacientes con enfermedad de Alzheimer.

En cuanto a la clasificación CN vs. EMCI, se aprecia en la Figura 30 como los resultados son en general peores, como cabía de esperar tras la visualización de texturas 2D (Apartado 5.1.1).

A su vez, como era de esperar, el área calculada bajo la curva ROC se corresponde con uno de los AUC más altos calculado (0,77), concretamente el tercero más alto, por lo que se puede afirmar que la curva ROC ha sido fiable, hecho que se observa en su forma (Figura 32) aunque no del todo determinante en la valoración de los resultados, pues el mayor AUC encontrado aparece para esta misma región de interés (ROI_2_I_2D) aunque en la clasificación CN vs. AD y con un *accuracy* del 76%.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

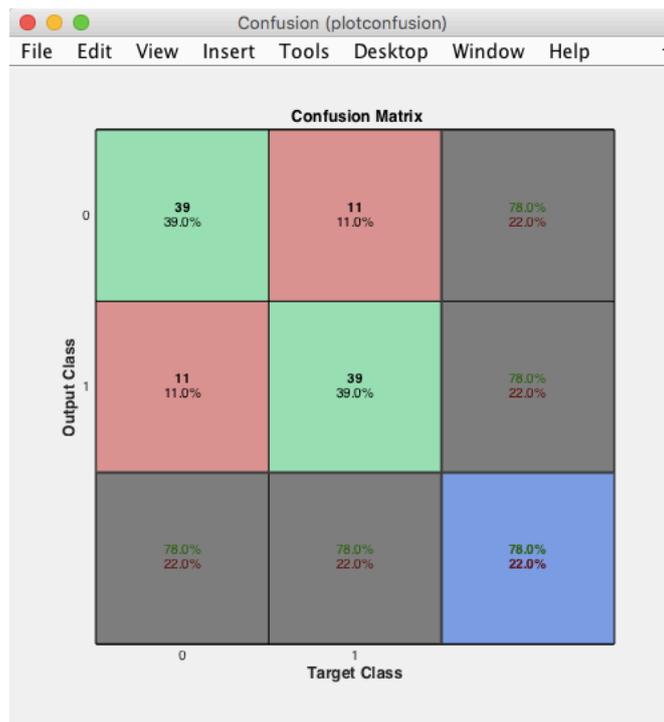


Figura 31. Matriz de confusión para la clasificación EMCI vs. AD de la ROI_2_I_2D. La clase EMCI aparece representada como 0 y AD como 1.

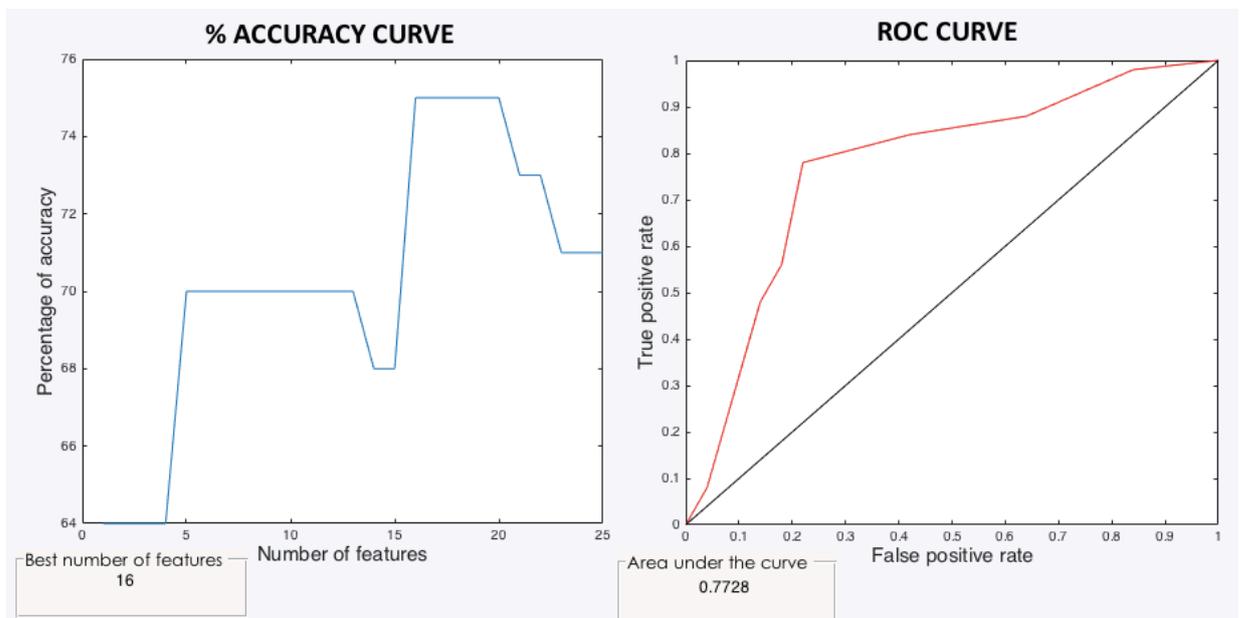


Figura 32. Curva *profile* y curva ROC para la clasificación EMCI vs. AD de la ROI_2_I_2D.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Si analizamos estos mismos resultados desde el punto de vista del tipo de clasificación en función de las clases, se observa cómo la clasificación binaria destaca frente a la *multiclass*, donde para esta última los resultados no han sido especialmente destacables. Conceptualmente la clasificación multiclase puede entenderse como una serie de clasificaciones binarias combinadas entre sí. En la Tabla 5 se aprecia cómo, para esta clasificación, el máximo porcentaje de *accuracy* obtenido es de 54%. Estos resultados tan bajos pueden deberse a la influencia de la clasificación de las clases CN vs. EMCI que, al producir también una mala clasificación binaria, contribuye negativamente al resultado.

Este hecho es apreciable en la Figura 33, donde se representa la matriz de confusión para la clasificación multiclase de la misma región de interés que se estudió anteriormente, la ROI_2_I_2D. Se observa como la clasificación CN vs. EMCI destaca por su alta contribución negativa en los resultados (11 pacientes control fueron clasificados como deterioro cognitivo leve temprano y 22 EMCI fueron clasificados como CN).



Figura 33. Matriz de confusión para la clasificación multiclase de ROI_2_I_2D.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Por otra parte, si se analizan los resultados desde el punto de vista del radio empleado para llevar a cabo la segmentación del hipocampo, se observa como parece que la región de interés del hipocampo izquierdo segmentada con un radio de 5 píxeles (ROI_2_I_2D) es la que, en general, ha experimentado resultados más notables exceptuando la clasificación CN vs. EMCI (ver Figura 30). Mientras que las regiones segmentadas con un radio de 3 píxeles tanto para el hipocampo izquierdo como para el derecho han sido las que poseen menores valores de *accuracy* (ROI_1_D_2D y ROI_1_I_2D). Este hecho puede estar relacionado con que, al emplear circunferencias de mayor radio para la segmentación, se capta con mayor probabilidad la totalidad de la lesión del hipocampo causada por los pacientes enfermos de Alzheimer, lo que conlleva una mejor diferenciación de los datos y, lo que explicaría que, con una segmentación circular de un círculo de 3 píxeles de radio se obtuvieran resultados más bajos. A su vez, se puede afirmar que la mejor clasificación se obtiene para las regiones de interés de radio 5 píxeles, tanto para el hipocampo izquierdo como para el hipocampo derecho (Figura 30), este hecho se explicaría siguiendo el mismo razonamiento: las regiones de 3 píxeles son insuficientes para captar la totalidad de la lesión, mientras que las de 8 píxeles captan el hipocampo y, a su vez, zonas circundantes a este, lo que provoca la aparición de cierto ruido en los datos finales y que, como consecuencia, sea la región de 5 píxeles la que logra los mejores resultados. Además, cabe resaltar que, tal y como se estudió durante la introducción de este trabajo, el hipocampo es una estructura tridimensional, con lo que, a pesar de aumentar el radio de la circunferencia para llevar a cabo la segmentación, no es posible captar la estructura completa de la lesión hipocampal pues, para ello, al menos se requeriría de una segmentación esférica (tridimensional).

5.3.2 Análisis de texturas 3D

A continuación, se hará una valoración de los resultados del análisis de texturas extraídas a partir de una segmentación esférica del hipocampo derecho e izquierdo para radios de 3, 5 y 8 píxeles respectivamente. Se seguirá la misma metodología de análisis que se empleó para las texturas 2D, evaluando principalmente el porcentaje de *accuracy* de cada una de las 6 regiones de interés, prevaleciendo siempre aquella que logre la máxima exactitud para el mínimo número de características. Una vez más, los resultados que se muestran a continuación se corresponden con los mejores resultados para cada clasificación, obtenidos a partir del método de validación cruzada (Tabla 6).

A su vez, para facilitar su análisis, estos se visualizan mediante un diagrama de barras en la Figura 34.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Clases	Región de interés	Método de clasificación	Tipo de distancia	Nº vecinos, <i>Box const.</i> o <i>splits</i>	Método de ponderación	Accuracy (%)	Nº de características	AUC
Multiclass	ROI_1_D_3D	SVM	Linear	100	P-value	42,68	7	-
	ROI_1_I_3D	KNN	Cityblock	3	P-value	44,67	23	-
	ROI_2_D_3D	SVM	Linear	1	Fisher	52,67	17	-
	ROI_2_I_3D	SVM	Linear	1	Fisher	55,33	12	-
	ROI_3_D_3D	KNN	Euclidean	3	P-value	51,33	20	-
	ROI_3_I_3D	SDT	-	5	P-value	52,00	4	-
CN vs. AD	ROI_1_D_3D	SVM	Linear	100	Fisher	61,00	3	0.65
	ROI_1_I_3D	SVM	Linear	1	Fisher	63,00	19	0.69
	ROI_2_D_3D	KNN	Euclidean	3	P_value	75,00	18	0.77
	ROI_2_I_3D	KNN	Cityblock	7	P-value	76,00	23	0.79
	ROI_3_D_3D	SVM	Linear	1	ReliefF	72,00	12	0.78
	ROI_3_I_3D	KNN	Euclidean	7	ReliefF	77,00	16	0.76
CN vs. EMCI	ROI_1_D_3D	SVM	Polynomial	1	P-value	60,00	1	0.63
	ROI_1_I_3D	SVM	Linear	1	ReliefF	60,00	2	0.61
	ROI_2_D_3D	SDT	-	3	ReliefF	67,00	2	0.62
	ROI_2_I_3D	SDT	-	5	Fisher	63,00	5	0.63
	ROI_3_D_3D	SVM	Polynomial	1	P-value	63,00	6	0.63
	ROI_3_I_3D	SVM	Polynomial	1	Fisher	61,00	22	0.62
EMCI vs. AD	ROI_1_D_3D	SVM	Gaussian	1	Fisher	50,00	24	0.50
	ROI_1_I_3D	SVM	Linear	100	P-value	72,00	10	0.69
	ROI_2_D_3D	SVM	Polynomial	100	Fisher	71,00	8	0.71
	ROI_2_I_3D	SVM	Linear	100	ReliefF	77,00	10	0.79
	ROI_3_D_3D	SDT	-	3	Fisher	70,00	16	0.69
	ROI_3_I_3D	SVM	Linear	100	Fisher	80,00	17	0.82

Tabla 6. Resultados para regiones de interés de 3 dimensiones.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

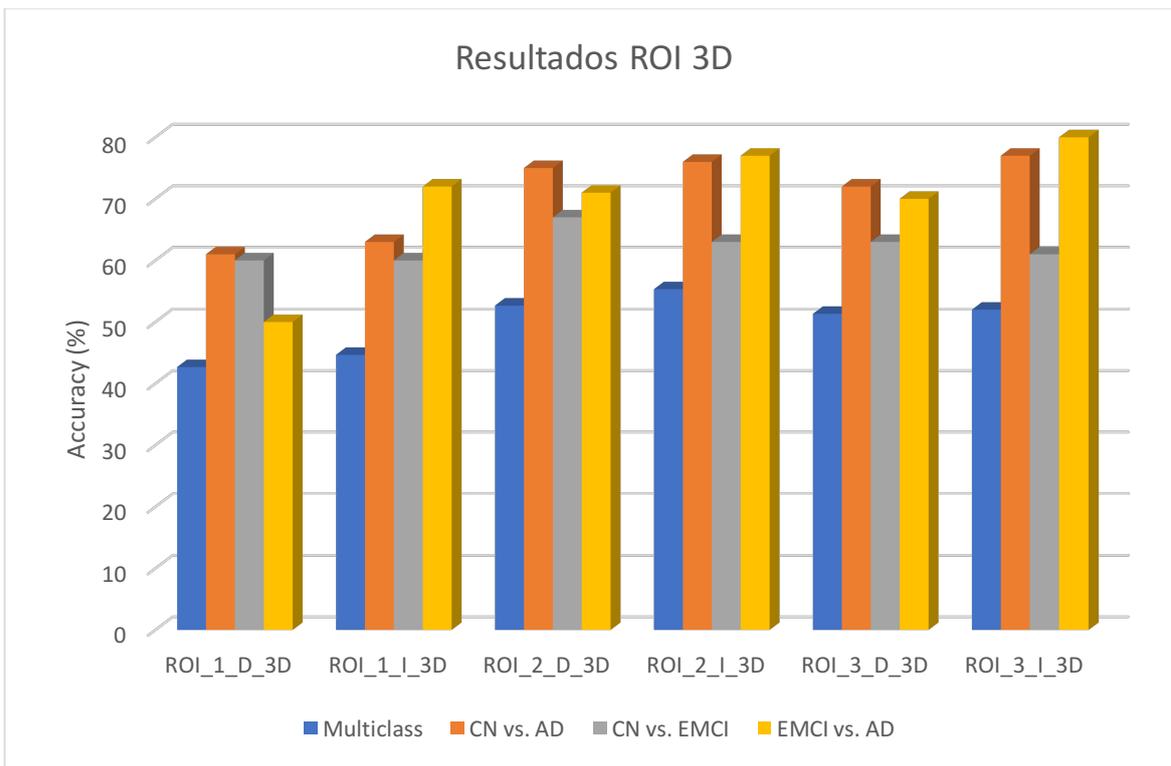


Figura 34. Visualización resultados de regiones de interés de 2 dimensiones.

Tal y como ocurría en la clasificación 2D, los peores resultados coinciden para la clasificación multiclase de las regiones ROI_1_D y ROI_1_I, concretamente el hipocampo derecho segmentado con una esfera de 3 píxeles de radio (ROI_1_D) obtiene la clasificación multiclase más baja. Principalmente, estos resultados pueden deberse a la elevada contribución negativa de la clasificación EMCI vs. AD, hecho que se ve reflejado en la matriz de confusión de la Figura 34, donde, a partir de una muestra de 150 pacientes, un total de 14 pacientes con deterioro cognitivo leve fueron clasificados con Alzheimer y 20 sujetos AD fueron clasificados como EMCI, es decir, un 22,6% del error de la clasificación multiclase (57,3%) se corresponde sólo con el error generado por la clasificación EMCI vs. AD.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

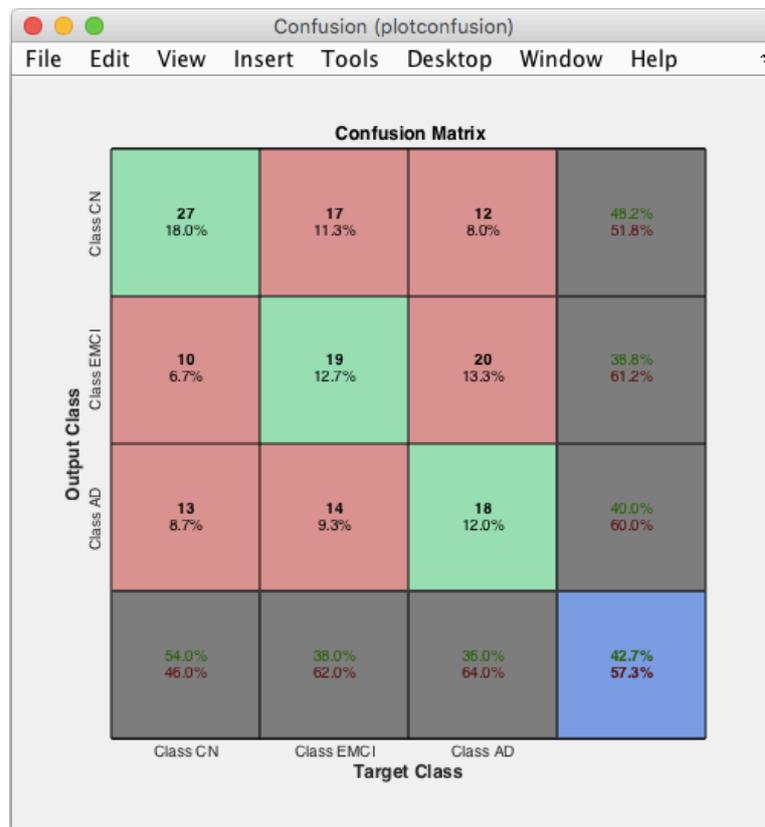


Figura 35. Matriz de confusión para la clasificación multiclase de ROI_1_D_3D.

Por otro lado, en general, los valores de exactitud más altos se vuelven a obtener para la clasificación EMCI vs. AD. Concretamente, en la región ROI_3_I_3D, se obtiene un 80% de *accuracy* para un modelo elaborado mediante la técnica de SVM lineal, con un *box constraint* de 100 y 17 de las 25 características, esta vez ordenadas mediante el ranking de Fisher.

Para este último modelo de clasificación se obtienen 41 muestras etiquetadas correctamente como pacientes con deterioro cognitivo leve temprano y 39 clasificadas de forma correcta como pacientes con enfermedad de Alzheimer, con una precisión del 78,8% y del 81,2% y con un ratio de verdaderos positivos del 82% y del 78% respectivamente. Únicamente 9 pacientes EMCI fueron incorrectamente clasificados como AD y 11 AD fueron clasificados incorrectamente como EMCI, lo que hace una tasa general de error del 20%, tal y como se observa en la Figura 36.

Se trata del porcentaje de *accuracy* más alto visto durante todo el estudio, y, como cabe esperar, se corresponde con el AUC más alto para la clasificación de texturas 3D, concretamente de 0,82. A pesar de que no se trata de una clasificación perfecta, estos son unos buenos resultados, hecho que se ve reflejado en la representación de la curva ROC (Figura 37), donde se observa cómo el área debajo de la curva tiende a aproximarse a 1.

A su vez, cabe destacar que en la clasificación CN vs. AD se consiguen también resultados destacables, llegando a alcanzar porcentajes de *accuracy* del 77% para, de nuevo, el caso de la ROI_3_I_3D. Por otra parte, una vez más los peores resultados para una clasificación binaria se observan en los grupos CN vs. EMCI.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

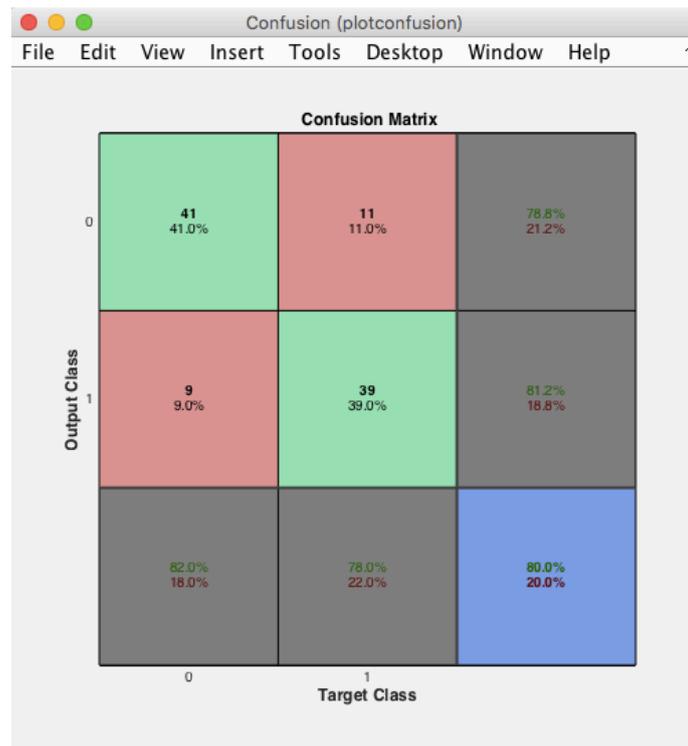


Figura 36. Matriz de confusión para la clasificación EMCI vs. AD de la ROI_3_I_3D. La clase EMCI aparece representada como 0 y la clase AD como 1

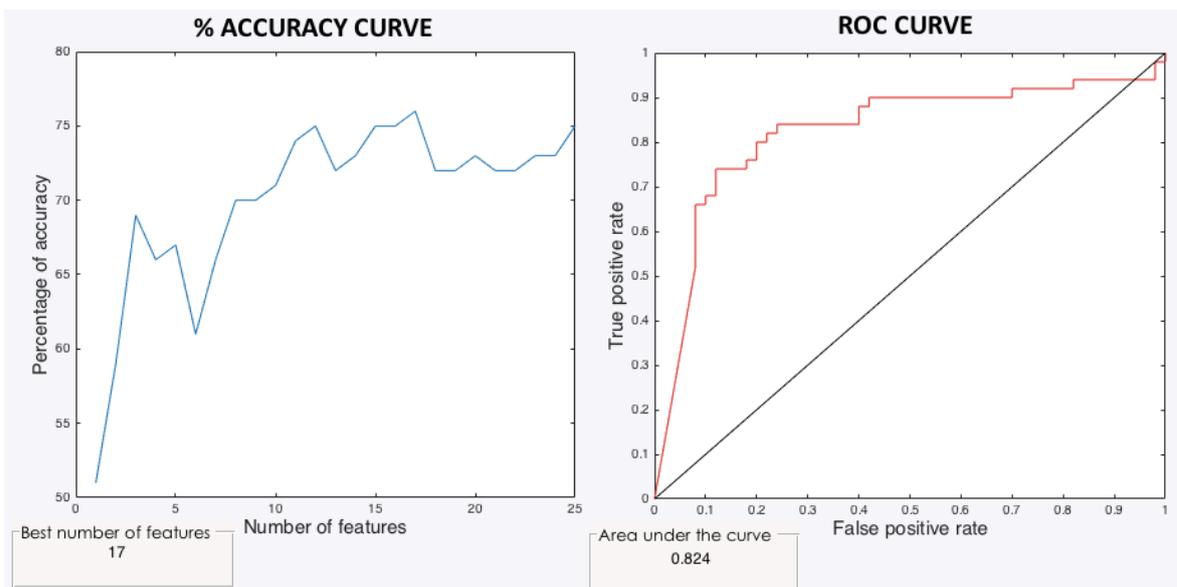


Figura 37. Curva *profile* y curva ROC para la clasificación EMCI vs. AD de la ROI_3_I_3D.

5.3.3 Comparación análisis 2D y 3D

En general, si observamos los correspondientes diagramas de barras para la clasificación en dos y tres dimensiones (Figura 30 y Figura 34, respectivamente), se aprecia como los resultados son ligeramente más altos para la segmentación 3D. Esta mejora se hace sobre todo notable en las regiones de interés de 8 píxeles (ROI_3_D y ROI_3_I), donde los porcentajes de precisión aumentan con respecto a los de texturas 2D para toda clasificación menos para CN vs. EMCI del hipocampo izquierdo, donde se experimenta una breve disminución (ver Figura 38).

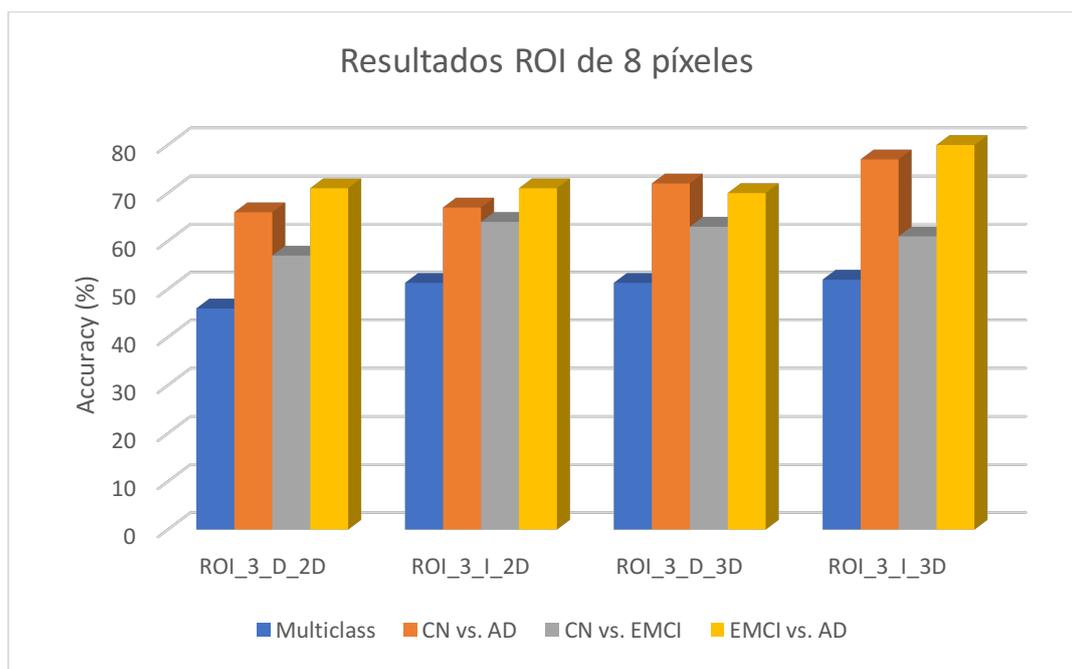


Figura 38. Comparación de los resultados 2D y 3D para una ROI de 8 píxeles.

Como previamente se ha deducido, puede existir una cierta relación entre el tamaño de la región de interés segmentada y los resultados obtenidos. A mayor región de interés, existe mayor probabilidad de segmentar completamente el hipocampo, sin embargo, el hecho de trabajar con regiones de interés circulares o esféricas, provoca que esta segmentación no sea del todo fiel a la forma de esta estructura cerebral, como consecuencia, al escoger regiones amplias es probable que se capten tejidos circundantes al hipocampo que puedan crear ruido en los resultados, como se dedujo para el caso de las estructuras en 2D. Sin embargo, para las texturas en tres dimensiones segmentadas con ROIs esféricas, se toman datos del volumen completo del hipocampo que, según se pudo observar durante la introducción, consta de una forma asimétrica en volumen. De este modo, aunque puedan recogerse a su vez datos de las regiones circundantes, se garantizará la toma de datos de su totalidad y, como consecuencia, el ruido causado por el resto de estructuras será menor.

5.4 Estudio de los métodos

Finalmente, se analizará cuáles han sido los métodos de clasificación para los cuales se han obtenido los mejores resultados con el fin de reconocer algún tipo de tendencia o relación entre los modelos de clasificación elaborados y el mejor porcentaje de precisión.

En general, si observamos con detalle los porcentajes de *accuracy* obtenidos (Tablas 4 y 5), el método de clasificación más empleado ha sido el *Support Vector Machine*. SVM ha destacado en la obtención de resultados tanto para texturas 2D como para 3D, en las Figuras 38 y 39 se observa cuál ha sido la frecuencia con la cual se ha empleado cada función, concretamente se aprecia como destaca la presencia de la función lineal para el cálculo del *kernel*. Un hecho a destacar reside en que, para la clasificación multiclase 2D, la totalidad de los resultados 2D se han calculado con modelos de clasificación SVM.

Si se analizan estos datos desde un punto de vista más genérico, concretamente en 31 de los 48 resultados se ha empleado SVM, lo que se traduce en que en un 64,58% de los modelos resultantes han sido calculados con *Support Vector Machine*. A su vez, en 12 ocasiones se ha obtenido el mejor modelo empleando el método de KNN y en 5 el de SDT.

Tal y como se aprecia en las Figuras 38 y 39 los métodos KNN y SDT se emplean con menos frecuencia para elaborar modelos de clasificación, tanto en texturas bidimensionales como tridimensionales. Para calcular la distancia al vecino más próximo (KNN) en el caso de las muestras de texturas 2D no se ha empleado en ninguna ocasión la distancia *Cityblock*; la distancia euclídea, por otro lado, ha sido empleada en seis ocasiones y en dos la de *Chebyshev*. Nuevamente, para las texturas 3D, la distancia más utilizada en KNN ha sido la euclídea (en tres ocasiones) y en este caso, la *Cityblock* en dos.

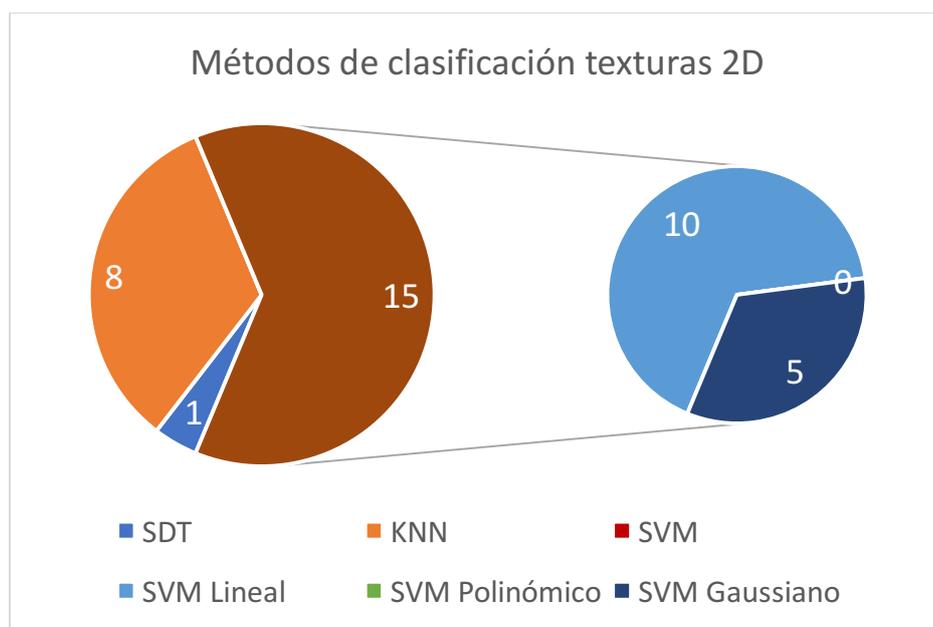


Figura 39. Frecuencia de uso de métodos de clasificación para resultados 2D.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

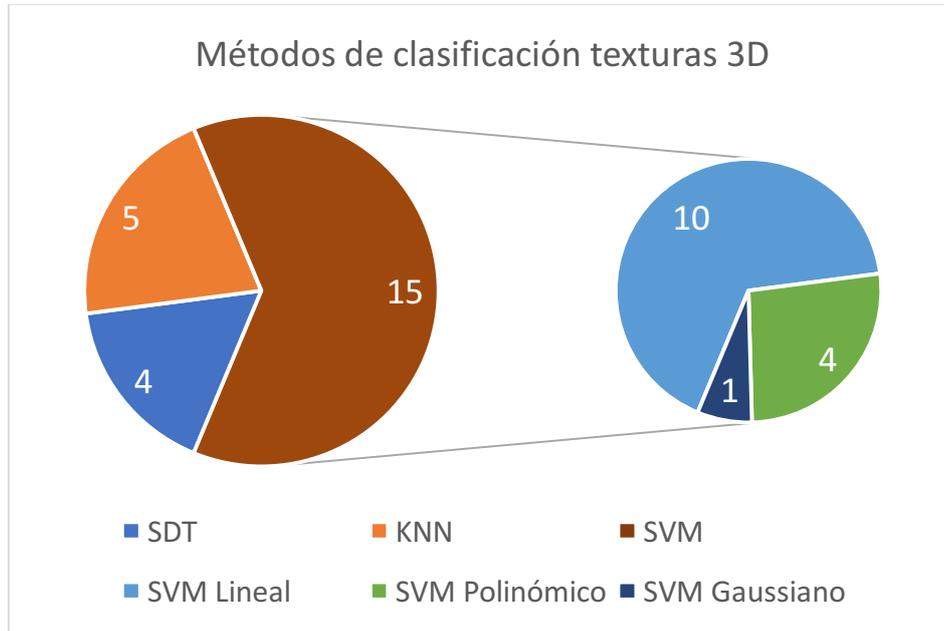


Figura 40. Frecuencia de uso de métodos de clasificación para resultados 3D.

Por último, en cuanto a los métodos de selección de características empleados para elaborar el ranking de las mejores características para llevar a cabo la clasificación, destaca especialmente el método del p -valor para texturas 2D (utilizado en 14 de las 24 pruebas) y el de Fisher para 3D (empleado en 10 de las 24 pruebas). La distribución de estos métodos en los resultados se observa en la Figura 41.

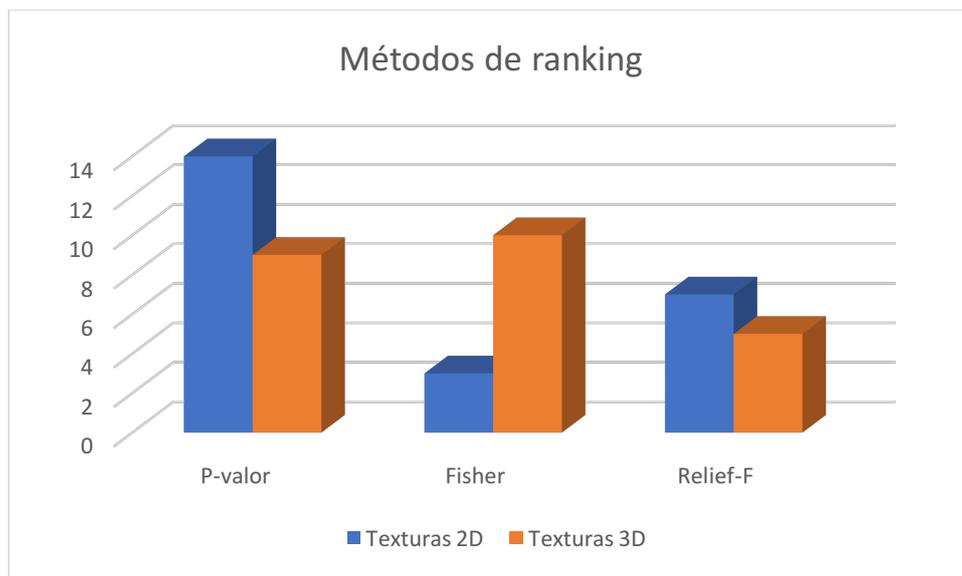


Figura 41. Frecuencia de uso de métodos de ranking para resultados 3D.

5.5 Discusión con otros estudios

Con el fin de contrastar los resultados obtenidos a lo largo de este estudio con razonamientos o procedimientos de otros autores para poder así extraer posteriormente conclusiones, se compararán algunos de los resultados con el conocimiento previo del tema estudiado.

En primer lugar, se observó cómo los peores resultados se obtuvieron en el caso de la región de interés más pequeña (3 píxeles), tanto para texturas bidimensionales como tridimensionales, tal y como ya se comentó, esto puede estar relacionado con que no se logre captar la totalidad de la lesión empleando ROIs de este tamaño. Determinados estudios confirman que regiones de interés pequeñas se encuentran relacionadas con malos resultados, hecho que concuerda con el razonamiento anteriormente explicado (Zhang *et al.*, 2012).

Además, los resultados sugieren que clasificación mejora en el caso del uso de texturas tridimensionales, para las cuales se llegó a obtener un *accuracy* del 80% en la clasificación usando 17 texturas. Jinhg Zhang ya confirmaba en su estudio (Zhang *et al.*, 2012) que las texturas 3D eran útiles para el diagnóstico de la enfermedad de Alzheimer, pero, además, afirmaba que la exactitud de la clasificación, es decir, el valor de *accuracy* no dependía del número de características empleadas, sino de cómo estas eran seleccionadas y combinadas entre sí. Para este trabajo en cuestión se obtuvieron distintos *rankings* que ordenaban aquellas texturas consideradas como idóneas de acuerdo a entrenar nuestro modelo de clasificación y, en muchas ocasiones, los mejores resultados no necesitaron de las 25 texturas extraídas inicialmente. Por ejemplo, para obtener el mejor *accuracy* de todo el estudio sólo se necesitaron 17 de esas 25 texturas ordenadas según el método de Fisher.

Por otra parte, se observó que el método de selección de características más empleado fue el *Support Vector Machine* (SVM), frente a *K-Nearest Neighbors* (KNN) y *Decision Trees* (DT). A pesar de haber probado con funciones polinómicas o gaussianas, el enfoque lineal fue el que logró mejores resultados con diferencia. Durante uno de los estudios de la clasificación de IMR para la detección de la enfermedad de Alzheimer (Klöppel *et al.*, 2008) se probó como el rendimiento de los núcleos no lineales para SVM no mejoraba la clasificación, sugiriendo así que el enfoque lineal era válido y adecuado para llevarla a cabo, hecho que se ha corroborado durante este trabajo.

6. CONCLUSIONES

El *machine learning* es una herramienta muy potente para la clasificación de datos que ofrece una infinidad de combinaciones para crear un modelo que se ajuste a las necesidades de cada clasificación. Esta idea ofrece muchas ventajas, sin embargo, también cuenta con el inconveniente de que, el análisis de datos puede resultar una tarea compleja debido a la gran cantidad de opciones y modelos disponibles, por lo que elaborar una herramienta clasificadora para encontrar parámetros significativos dentro de estos grupos de datos puede no resultar tarea fácil. No obstante, la detección precoz del Alzheimer empleando métodos de aprendizaje máquina es una temática que ha generado gran interés entre los investigadores en los últimos años y se ha producido una enorme cantidad de literatura al respecto, aunque actualmente no se haya logrado modificar el curso de esta enfermedad.

A lo largo del desarrollo de este trabajo se han obtenido varias conclusiones. En primer lugar, se ha observado que el tamaño de la región de interés para extraer texturas, tanto 2D como 3D es un parámetro determinante en la clasificación, tal y como se ha discutido, pues significa tomar en su totalidad la región afectada por la enfermedad, en este caso el hipocampo. Definir la posición del círculo o la esfera para tratar de captar toda la información necesaria para analizar el tejido es algo complejo y, como consecuencia, pueden perderse detalles útiles para la clasificación. A su vez, varios estudios confirman que la definición manual de la ROI de forma que esta se ajuste a los límites del hipocampo es un punto muy favorable en la clasificación (Larroza, Bodí y Moratal, 2016), en este estudio se han extraído texturas 2D y 3D a partir de regiones circulares y esféricas respectivamente, es por ello que quizás los resultados no sean del todo concluyentes, aunque se ha observado una mejora en la clasificación a partir de la extracción de texturas 3D que termina de confirmar cómo mejora el estudio al estudiar la zona afectada en volumen.

La elección del clasificador y del método de *feature selection* influye notablemente en el resultado final. Se han logrado resultados especialmente destacables para la clasificación CN vs. AD y EMCI vs. AD, donde, tanto los porcentajes de *accuracy* como el valor del área debajo de la curva ROC (AUC) superan el 75%, mientras que, en el caso de la clasificación CN vs. EMCI y la clasificación multiclase los resultados no han sido especialmente satisfactorios. Asimismo, cabe recordar que, para elaborar el clasificador únicamente se ha empleado una muestra de 150 pacientes. Con el fin de lograr una técnica de clasificación robusta es necesario entrenar el sistema empleando un mayor número de datos, de este modo, análisis que realizaría el modelo sería más exhaustivo y veraz. A pesar del pequeño tamaño muestral, el hecho de haber empleado la validación cruzada ha permitido obtener resultados que se adapten fielmente a la realidad.

Resulta de elevada importancia continuar con el desarrollo de instrumentos de evaluación de imágenes y datos, así como de análisis de biomarcadores, con el fin de evaluar las etapas asintomáticas de las enfermedades neurodegenerativas y detectar individuos que se encuentren en riesgo potencial. Sin duda, la inteligencia artificial y, en concreto, las técnicas de aprendizaje automático desarrollan un papel indispensable en este ámbito, por lo que resulta vital continuar investigando y mejorando modelos que permitan la predicción y clasificación de pacientes con patologías como el Alzheimer.

7. LÍNEAS FUTURAS

En primer lugar, como posible línea futura a desarrollar en este trabajo, aumentar el tamaño de la muestra de datos podría mejorar la capacidad de predicción del modelo, por lo que, en un futuro, se podría tratar de ampliar el número de pacientes estudiados con el fin de mejorar la robustez del modelo y poder llegar a alguna conclusión que pueda generalizarse.

Partiendo de la herramienta creada, se podría tratar de consolidar una herramienta más compleja que incorpore otros métodos de clasificación, *feature selection* y validación. De este modo, se dispondría de numerosas alternativas que permitan la adaptación del modelo al problema en cuestión, logrando poner al alcance del usuario un amplio abanico de recursos que le permitan modificar y adaptar el modelo a las condiciones requeridas.

A su vez, se podría mejorar el estudio realizando una segmentación manual que se ajuste a los límites de las regiones de interés, de modo que se garantizase que se capta la totalidad de la lesión, minimizando la captación de tejidos circundantes al hipocampo que puedan modificar los valores de los resultados y disminuyendo la probabilidad de perder información útil para el modelo.

Además, con el fin de aumentar la precisión de *Classifiers* y estudiar en mayor detalle las imágenes de resonancia magnética de cada paciente, el siguiente punto a desarrollar podría ser la ampliación de los métodos de análisis de texturas usando un método descriptor visual, concretamente *local binary pattern* (LBP). Este método divide cada imagen en un número determinado de celdas y adquiere el histograma para cada una de ellas, normalizando y concatenando los valores obtenidos para cada celda obtiene un vector de características (*feature vector*) que podría ser utilizado para la clasificación del paciente. Se trata de un método clave en la clasificación de texturas, por lo que podría llegar a desarrollar un importante papel en la detección precoz de Alzheimer a partir de imágenes de resonancia magnética y un buen complemento para la herramienta desarrollada en este trabajo.

8. REFERENCIAS

- Aguilar Roblero, R. (2015) «El sistema de posicionamiento cerebral: Premio Nobel en Fisiología y Medicina 2014.», *Revista de la Facultad de Medicina de la UNAM*, 58.
- Ahmed, S. *et al.* (2008) «Predicting rapid clinical progression in amnesic mild cognitive impairment», *Dementia and Geriatric Cognitive Disorders*, 25(2), pp. 170-177. doi: 10.1159/000113014.
- Arlot, S. y A. Celisse (2009) «A survey of cross-validation procedures for model selection», 4, pp. 40-79. doi: 10.1214/09-SS054.
- Bishop, C. M. (2013) *Pattern Recognition and Machine Learning*, *Journal of Chemical Information and Modeling*. doi: 10.1117/1.2819119.
- Casella, G., S. Fienberg e I. Olkin (2006) *An Introduction to Statistical Learning, Design*. doi: 10.1016/j.peva.2007.06.006.
- Castellano, G. *et al.* (2004) «Texture analysis of medical images», *Clinical Radiology*, 59(12), pp. 1061-1069. doi: 10.1016/j.crad.2004.07.008.
- Chaplot, S., M. L. Patnaik, y N. R. Jagannathan (2006) «Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network», *Biomedical Signal Processing and Control*, 1(1), pp. 86-92. doi: 10.1016/j.bspc.2006.05.002.
- Classifier comparison* (fecha de acceso: 22-06-2018). Disponible en: http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html (Accedido: 22 de junio de 2018).
- Clinical Stages of Alzheimer's | Fisher Center for Alzheimer's Research Foundation* (fecha de acceso: 08-05-2018). Disponible en: <http://www.alzinfo.org/understand-alzheimers/clinical-stages-of-alzheimers/> (Accedido: 8 de mayo de 2018).
- Duda, R. O., P. E. Hart y D. G. Stork (2001) «Pattern Classification», *New York: John Wiley, Section*, p. 680. doi: 10.1007/BF01237942.
- Fisher, R. A. (1936) «The Use of Multiple Measurements in Taxonomic Problems», *Annals of Eugenics*, 7(2), pp. 179-188. doi: 10.1111/j.1469-1809.1936.tb02137.x.
- Freeborough, P. A. y N. C. Fox (1998) «MR Image Texture Analysis Applied to the Diagnosis and Tracking of Alzheimer's Disease», *IEEE Transactions on Medical Imaging*, 17(3), pp. 8-11. doi: 10.1109/42.712137.
- Gu, Q., Z. Li y J. Han (2012) «Generalized Fisher Score for Feature Selection». Disponible en: <http://arxiv.org/abs/1202.3725>.
- Hastie, T., R. Tibsharani y J. Friedman (2009) «Springer Series in Statistics The Elements of», *The Mathematical Intelligencer*, 27(2), pp. 83–85. doi: 10.1007/b94608.
- Jack, C. R. *et al.* (2008) «11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnesic mild cognitive impairment», *Brain*, 131(3), pp. 665-680. doi: 10.1093/brain/awm336.
- Jiménez Bonilla, J. F. y J. M. Carril Carril (2013) «La neuroimagen molecular en las demencias degenerativas», *Revista Espanola de Medicina Nuclear e Imagen Molecular*. SEMNIM, 32(5), pp. 301-309. doi: 10.1016/j.remnm.2013.06.012.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

- Klöppel, S. *et al.* (2008) «Automatic classification of MR scans in Alzheimer's disease», *Brain*, 131(3), pp. 681-689. doi: 10.1093/brain/awm319.
- Knopman, D. S. *et al.* (2001) «Practice parameter: Diagnosis of dementia (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology», *Neurology*, (56), pp. 1143-1153. doi: 10.1212/WNL.56.9.1143.
- Kuhn, M. y K. Johnson (2013) *Applied Predictive Modeling [Hardcover]*. doi: 10.1007/978-1-4614-6849-3.
- Larroza, A., Bodí, V. y Moratal, D. (2016) «Texture Analysis Magnetic Resonance Imaging : Considerations for Future Applications Review and Considerations for Future Applications», (5). doi: 10.5772/64641.
- Lemm, S. *et al.* (2011) «Introduction to machine learning for brain imaging», *NeuroImage*. Elsevier Inc., 56(2), pp. 387-399. doi: 10.1016/j.neuroimage.2010.11.004.
- López Gómez, C. (2017) «Análisis y evaluación de nuevos biomarcadores para la enfermedad de Alzheimer mediante un proceso radiomics basado en el análisis de texturas sobre imágenes de resonancia magnética». Disponible en: <https://riunet.upv.es:443/handle/10251/84937> (Accedido: 22 de junio de 2018).
- Martí-Climent, J. M. *et al.* (2010) «Neuroimagen: Fundamentos técnicos y prácticos», *Revista Española de Medicina Nuclear*, 29(4), pp. 189-210. doi: 10.1016/j.remn.2010.05.001.
- Molinuevo Guix, J. L. (2018) «Revista Española de Geriátría y Gerontología», 53(1), pp. 53-54. doi: 10.1016/j.regg.2017.01.003.
- Moratal, D. (2017) «Técnicas de Imágenes Biomédicas Tema 9. Imagen por resonancia magnética 9.1. Introducción a la imagen por RM», pp. 1-22.
- Petersen, R. C. *et al.* (2009) «Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization», *Neurology*, 74(3), pp. 201-209. doi: 10.1212/WNL.0b013e3181cb3e25.
- Pose, M. y F. Manes (2010) «Deterioro cognitivo leve», *Acta Neurológica Colombiana*, 26(3), pp. 7-12. doi: 10.4067/S0717-92272003000200005.
- Reisberg, B. *et al.* (1982) «The Global Deterioration Scale for Assessment of Primary Degenerative Dementia», *Am J Psychiatry*, 139(9).
- Robnik-Sikonja, M. e I. Kononenko (2003) «Theoretical and empirical analysis of Relief and ReliefF», *Machine Learning*, 53, pp. 23-69.
- Roffo, G. *et al.* (2017) «Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach». Disponible en: <http://arxiv.org/abs/1707.07538> (Accedido: 16 de mayo de 2018).
- Roffo, G. (2017) «Ranking to Learn and Learning to Rank: On the Role of Ranking in Pattern Recognition Applications». Disponible en: <http://arxiv.org/abs/1706.05933> (Accedido: 16 de mayo de 2018).
- Roffo, G. y S. Melzi (2017) «Ranking to Learn», en, pp. 19-35. doi: 10.1007/978-3-319-61461-8_2.
- Sosa-Ortiz, A. L., I. Acosta-Castillo y M. J. Prince (2012) «Epidemiology of Dementias and Alzheimer's Disease», *Archives of Medical Research*. Elsevier Inc, 43(8), pp. 600-608. doi: 10.1016/j.arcmed.2012.11.003.
- Squire, L. R. (1992) «Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory», *Journal of Cognitive Neuroscience*, 4(3), pp. 232-243. doi: 10.1162/jocn.1992.4.3.232.

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

Tang, J., S. Alelyani y H. Liu (2014) «Feature Selection for Classification: A Review», *Data Classification: Algorithms and Applications*, pp. 37-64. doi: 10.1.1.409.5195.

Vallières, M. *et al.* (2015) *A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities*, *Physics in Medicine and Biology*. doi: 10.1088/0031-9155/60/14/5471.

Valls-Pedret, C., J. L. Molinuevo y L. Rami (2010) «Diagnóstico precoz de la enfermedad de Alzheimer: fase podrómica y preclínica», *Revista de Neurología*, 51(8), pp. 471-480.

Wortmann, M. (2012) «Dementia: A global health priority - Highlights from an ADI and World Health Organization report», *Alzheimer's Research and Therapy*, 4(5), pp. 4-6. doi: 10.1186/alzrt143.

Zhang, J. *et al.* (2012) «3D texture analysis on MRI images of Alzheimer's disease», *Brain Imaging and Behavior*, 6 (1), pp. 61-69. doi: 10.1007/s11682-011-9142-3.

DOCUMENTO

PRESUPUESTO

**ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE
TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA
DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE
ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA**

AUTORA: CAROLINA GIMÉNEZ ARIAS

TUTOR: DAVID MORATAL PÉREZ

Centro de Biomateriales e Ingeniería Tisular, Universitat Politècnica de València

COTUTOR: RAFAEL ORTIZ RAMÓN

Centro de Biomateriales e Ingeniería Tisular, Universitat Politècnica de València

CURSO 2017-18

ÍNDICE PRESUPUESTO

1.	INTRODUCCIÓN	4
2.	CUADRO DE PRECIOS MANO DE OBRA	5
3.	CUADRO DE PRECIOS MATERIALES, EQUIPOS Y SOFTWARE.....	6
4.	CUADRO DE PRESUPUESTOS PARCIALES	7
4.1.	Planificación del proyecto y conocimiento del estado del arte	7
4.2.	Programación software.....	7
4.3.	Análisis de nuevos biomarcadores	7
4.4.	Redacción y defensa del TFG	7
5.	CUADRO DE PRECIOS UNITARIOS.....	8
6.	CUADRO DE PRECIOS DESCOMPUESTOS	9
7.	PRESUPUESTO DE EJECUCIÓN POR CONTRATA.....	11

1. INTRODUCCIÓN

A lo largo de este documento se presenta una valoración económica del trabajo realizado. Como principal objetivo se pretende realizar una evaluación de la inversión económica estimada que supondría la implantación de la aplicación desarrollada para el análisis de biomarcadores en la enfermedad de Alzheimer.

Para elaborar el documento de presupuesto se han considerado varios aspectos que cabe destacar. En primer lugar, se evaluará la mano de obra necesaria para llevar a cabo este trabajo, concretamente esta mano de obra se centra en la acción de un ingeniero biomédico *junior*. También se mencionarán los costes asociados a materiales como el ordenador portátil y las licencias de los programas empleados. Seguidamente, se realizará un estudio donde se desglosen los presupuestos parciales en un cuadro, ordenados según cuatro capítulos que resumen el procedimiento de actuación para la realización del proyecto:

Capítulo 1. Planificación del proyecto y conocimiento del estado del arte

Capítulo 2. Programación software

Capítulo 3. Análisis de nuevos biomarcadores

Capítulo 4. Redacción y defensa del TFG

A continuación, se evaluará detalladamente la relación de los precios de las unidades de obra en un cuadro de precios unitarios, seguido por un cuadro de precios descompuestos donde se detallan las cantidades de mano de obra. Por último, se evaluará el presupuesto de ejecución por contrata.

2. CUADRO DE PRECIOS MANO DE OBRA

Teniendo en cuenta que, según el BOE del Ministerio de Empleo y Seguridad Social publicado en enero de 2018, durante el año 2018 la cotización de un ingeniero oscila entre 1.199,10 y 3.751,20 € al mes, para calcular el coste por hora de un ingeniero biomédico *junior*, se estima un sueldo base de 1.500€ al mes. A su vez, considerando que el ingeniero biomédico junior percibe dos pagas extra de la misma cuantía que un salario mensual, el sueldo anual sería de 21.000€.

Para calcular las horas trabajadas por parte del ingeniero junior consideraremos una jornada laboral de 8 horas. Descontando 15 días festivos debidos a fiestas nacionales y 30 días hábiles debidos a período vacacional y, teniendo en cuenta que, en un año hay 52 fines de semana, 4 de los cuales pertenecen al período vacacional, nos quedan 224 días de trabajo efectivo al año.

Estos 224 días resultan un total de 1792 horas de trabajo efectivo al año. Dividiendo el coste total para la empresa (27.720€/año aproximadamente) entre este número de horas, se obtiene un coste por hora de 15,47€ aproximadamente.

Para calcular el coste total del ingeniero biomédico junior para la empresa en concepto de seguridad social, se tiene en cuenta que, según el Ministerio de Empleo y Seguridad Social las cotizaciones para las contingencias comunes de las que se hace cargo la empresa se corresponden con el 23,60%. Además, la empresa se hace cargo del 5,5% de seguro de desempleo en caso de contratación indefinida y de las contingencias de accidentes de trabajo y enfermedades profesionales (1,65% de IT/IMS). A su vez, se deberá tener en cuenta el 0,20% de FOGASA (Fondo de Garantía Salarial) y el 0,60% correspondiente a la Formación Profesional.

<i>Nº</i>	<i>Código</i>	<i>Denominación de la mano de obra</i>	<i>Precio</i>	<i>Horas</i>	<i>Total</i>
1.	MO.ING	Ingeniero biomédico <i>junior</i>	15,47€/h	450	6.961,5€
<i>Total mano de obra</i>					6.961,5€

3. CUADRO DE PRECIOS MATERIALES, EQUIPOS Y SOFTWARE

Para elaborar el cuadro de precios materiales, equipos y softwares, los precios de cada producto se han obtenidos de las páginas web de los respectivos fabricantes y el factor de amortización se ha calculado como el período de tiempo que comprende el proyecto.

Cabe destacar que, para la realización de este trabajo principalmente se ha empleado el software MATLAB R2015b, aunque, además se ha empleado Microsoft Office, concretamente Microsoft Word para la redacción de la memoria y el presupuesto, Microsoft Excel para cálculos con algunos datos (que se exportaban directamente desde MATLAB a una tabla Excel) y Microsoft PowerPoint para la preparación de la defensa. Además, para adjuntar a la memoria algunas de las imágenes de resonancia magnética empleadas en el estudio, ha sido necesaria la previa edición de estas, para ello se ha empleado la versión Pro del programa Photoscape X.

Nº	Código	Denominación del material	Precio	Cantidad	Factor amortización	Total
1.	MA.PC	Ordenador portátil MacBook Air 1,6 GHz Intel Core i5 / 128GB / Intel HD Graphics 6000 / 13.3"	1.105,59€	1 u	8/48	184,27€
2.	MA.OF	Microsoft Office Hogar y Estudiantes 2016 para Mac	149€	1 u	8/48	24,83€
3.	MA.MAT	MATLAB R2015b	2000€	1 u	8/48	333,33€
4.	MA.ED	PhotoScape X Pro – Photo Editor	43,99€	1 u	3/48	2,75€
5.	MA.IM	Impresión en blanco y negro del TFG	0,036€	93 u	--	3,35€
6.	MA.EN	Encuadernación del TFG	3,50€	1 u	--	3,50€
<i>Total materiales</i>						555,03€

4. CUADRO DE PRESUPUESTOS PARCIALES

4.1. Planificación del proyecto y conocimiento del estado del arte

Nº	Ud	Denominación	Cantidad	Precio	Total
4.1.1	h	Reuniones para presentar la temática del proyecto y planificar su desarrollo	2	15,47€	30,94€
4.1.2	u	Reuniones periódicas para evaluar el grado de desarrollo del proyecto	20	23,21€	464,20€
4.1.3	h	Revisión de la literatura científica sobre técnicas de aprendizaje máquina, enfermedad del Alzheimer, análisis de texturas y segmentación	26	30,94€	804,44€
<i>Total presupuesto parcial 4.1</i>					1299,58€

4.2. Programación software

Nº	Ud	Denominación	Cantidad	Precio	Total
4.2.1	h	Diseño interfaz gráfica	5	15,47€	77,35€
4.2.2	h	Programación código	180	15,47€	2784,60€
4.2.3	u	Instalación programas necesarios	1	638	638€
4.2.4	h	Depuración código	6	15,47€	92,98€
4.2.5	h	Pruebas	4	15,47€	61,88€
<i>Total del presupuesto parcial 4.2</i>					3654,81€

4.3. Análisis de nuevos biomarcadores

Nº	Ud	Denominación	Cantidad	Precio	Total
4.3.1	h	Segmentación de imágenes	5	15,47€	77,35€
4.3.2	h	Evaluación de resultados	15	15,47€	232,05€
<i>Total del presupuesto parcial 4.3</i>					309,40€

4.4. Redacción y defensa del TFG

Nº	Ud	Denominación	Cantidad	Precio	Total
4.4.1	h	Redacción del proyecto	70	15,47€	1082,90€
4.4.2	h	Preparación presentación	20	15,47€	309,40€
4.4.3	u	Impresión y encuadernación	1	6,85€	6,85€
4.4.4	h	Revisiones y correcciones	5	15,47€	77,35€
<i>Total presupuesto parcial 4.4</i>					1476,5€

5. CUADRO DE PRECIOS UNITARIOS

Nº	Ud	Denominación	Importe	
			En cifra (€)	En letra (Euros)
1. Planificación del proyecto y conocimiento del estado del arte				
1.1	h	Reuniones para presentar la temática del proyecto y planificar su desarrollo	15,47	Quince euros con cuarenta y siete céntimos
1.2	h	Reuniones periódicas para evaluar el grado de desarrollo del proyecto	23,21	Veintitrés euros con veintiún céntimos
1.3	h	Revisión de la literatura científica sobre técnicas de aprendizaje máquina, enfermedad del Alzheimer, análisis de texturas y segmentación	30,94	Treinta euros con noventa y cuatro céntimos
2. Programación del software				
2.1	h	Diseño interfaz gráfica	15,47	Quince euros con cuarenta y siete céntimos
2.2	u	Instalación programa	638	Seiscientos treinta y ocho euros
2.3	h	Programación código	15,47	Quince euros con cuarenta y siete céntimos
2.4	h	Depuración código	15,47	Quince euros con cuarenta y siete céntimos
3. Análisis de nuevos biomarcadores				
3.1	h	Segmentación de imágenes	15,47	Quince euros con cuarenta y siete céntimos
3.2	h	Evaluación de resultados	15,47	Quince euros con cuarenta y siete céntimos
4. Redacción y defensa del TFG				
4.1	h	Redacción del proyecto	15,47	Quince euros con cuarenta y siete céntimos
4.2	h	Preparación presentación	15,47	Quince euros con cuarenta y siete céntimos
4.3	u	Impresión y encuadernación	6,85	Seis euros con ochenta y cinco céntimos
4.4	h	Revisiones y correcciones	15,47	Quince euros con cuarenta y siete céntimos

6. CUADRO DE PRECIOS DESCOMPUESTOS

Nº	Ud	Denominación			Total
1. Planificación del proyecto y conocimiento del estado del arte					
1.1	h	Reuniones para presentar la temática del proyecto y planificar su desarrollo			
		MO.ING	1 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					15,47€
				Precio total por h:	15,47€
1.2	h	Reuniones periódicas para evaluar el grado de desarrollo del proyecto			
		MO.ING	1,5 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					23,21€
				Precio total por u:	23,21€
1.3	h	Revisión de la literatura científica sobre técnicas de aprendizaje máquina, enfermedad del Alzheimer, análisis de texturas y segmentación			
		MO.ING	2 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					30,94€
				Precio total por h:	30,94€
2. Programación del software					
2.1	h	Diseño interfaz gráfica			
		MO. ING	1 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					15,47€
				Precio total por h:	15,47€
2.2	u	Instalación programas necesarios			
		MO.ING	6 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					92,82€
		MA. PC	1 u	Ordenador portátil MacBook Air 1,6 GHz Intel Core i5 / 128GB / Intel HD Graphics 6000 / 13.3"	1.105,59€ /u 8/48
					184,27€
		MA. MAT	1 u	MATLAB R2015b	2000€/u 8/48
					333,33€
		MA. OF	1 u	Microsoft Office Hogar y Estudiantes 2016 para Mac	149€/u 8/48
					24,83€
		MA. ED	1 u	PhotoScape X Pro – Photo Editor	43,99€/u 3/48
					2,75€
				Precio total por u:	638€
2.3	h	Programación código			
		MO.ING	1,5 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					23,21€
				Precio total por h:	23,21€
2.4	h	Depuración código			
		MO.ING	1 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					15,47€
				Precio total por h:	15,47€
3. Análisis de nuevos biomarcadores					
3.1	h	Segmentación de imágenes			
		MO.ING	1 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					15,47€
				Precio total por h:	15,47€
3.2	h	Evaluación de resultados			
		MO.ING	1 h	Ingeniero biomédico <i>junior</i>	15,47 €/h
					15,47€
				Precio total por h:	15,47€

ESTUDIO DE PARÁMETROS DE TEXTURA MEDIANTE TÉCNICAS DE APRENDIZAJE MÁQUINA PARA LA DETECCIÓN TEMPRANA DE LA ENFERMEDAD DE ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA

4. Redacción y defensa del TFG

4.1	h	Redacción del proyecto				
		MO.ING	4 h	Ingeniero biomédico <i>junior</i>	15,47 €/h	15,47€
					Precio total por h:	15,47€
4.2	h	Preparación presentación				
		MO.ING	4 h	Ingeniero biomédico <i>junior</i>	15,47 €/h	15,47€
					Precio total por h:	15,47€
4.3	u	Impresión y encuadernación				
		MA.IM	93 u	Impresión en blanco y negro del TFG	0,036€/ud	3,35€
		MA.EN	1 u	Encuadernación del TFG	3,50€/ud	3,50€
					Precio total por u:	6,85€
4.4	h	Revisiones y correcciones				
		MO.ING	4 h	Ingeniero biomédico <i>junior</i>	15,47 €/h	15,47€
					Precio total por h:	15,47€

7. PRESUPUESTO DE EJECUCIÓN POR CONTRATA

Para obtener el valor del presupuesto de ejecución por contrata de este proyecto, se ha obtenido el presupuesto de ejecución material como la suma de los presupuestos parciales. A continuación, se ha incrementado dicho valor en un 13% en concepto de gastos generales y en un 6% en concepto de beneficio industrial, por último, se le ha aplicado el 21% correspondiente al impuesto de venta adquirido (IVA). Cabe destacar que, el IVA sólo se aplica al final debido a que los costes de los recursos y tareas de distintos los capítulos del presupuesto no lo incluían.

	<i>Importe</i>
Capítulo 1. Planificación del proyecto y conocimiento del estado del arte	1299,58€
Capítulo 2. Programación software	3654,81€
Capítulo 3. Análisis de nuevos biomarcadores	309,40€
Capítulo 4. Redacción y defensa del TFG	1476,5€
Presupuesto de ejecución material	6.740,29€
Gastos generales (13%)	876,24€
Beneficio industrial (6%)	404,42€
Suma	8.020,95€
IVA (21%)	1.684,40€
Presupuesto de ejecución por contrata	9.705,35€

El presupuesto de ejecución por contrata asciende a la cantidad de **nueve mil setecientos cinco euros con treinta y cinco céntimos**.